

DOCUMENT RESUME

ED 102 162

95

SP 008 948

AUTHOR Houston, W. Robert, Ed.
 TITLE Competency Assessment, Research, and Evaluation.
 INSTITUTION Multi-State Consortium on Performance-Based Teacher Education, Albany, N.Y.
 SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.; Office of Education (DHEW), Washington, D.C.
 PUB DATE Mar 74
 NOTE 273p.; Report of a National Conference (University of Houston, Houston, Texas, March 12-15, 1974)
 AVAILABLE FROM American Association of Colleges for Teacher Education, Suite 610, 1 Dupont Circle, Washington, D.C. 20036 (\$3.00); National Dissemination Center for Performance Based Education, 150 Marshall St., Syracuse University, Syracuse, New York 13210 (\$3.00)
 EDRS PRICE MF-\$0.76 HC-\$13.32 PLUS POSTAGE
 DESCRIPTORS *Conceptual Schemes; *Educational Research; Organizational Change; *Performance Based Teacher Education; Professional Training; *Program Effectiveness; Social Factors; Teaching Techniques

ABSTRACT

This book is a compilation of papers that consider the current state of the art of competency/performance-based teacher education. Organized into six parts, the first part of this book reports three frames for conceptualizing research in competency-based education. The keynote address outlines five major areas for research: assessment of professional competency, program effectiveness, effectiveness of instructional strategies, social context, and institutional change. Parts two through five focus on these areas of research. (Editor/PD)

COMPETENCY ASSESSMENT, RESEARCH, AND EVALUATION

A Report of a National Conference
March 12-15, 1974

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
WASHINGTON, D.C. 20540

3971 111 111 111

BEST COPY AVAILABLE

This publication is being distributed by the Multi-State Consortium on Performance-Based Teacher Education as a function of one of its major objectives — “the dissemination and communication of information about performance-based teacher education.”

The opinions expressed herein should not be construed as representing the opinions of the United States Government, the Multi-State Consortium on Performance-Based Teacher Education or any of the states which are members of the Consortium, or the National Dissemination Center for Performance-Based Education, Syracuse University.

**COMPETENCY ASSESSMENT,
RESEARCH, AND EVALUATION**

A Report of a
National Conference
University of Houston
March 12-15, 1974

Sponsored by

The Multi-State Consortium on Performance Based Teacher Education
The National Commission on Performance Based Education
The National Consortium of CBE Centers
The National Institute of Education, Program on Teaching and Curriculum
The Committee on Performance Based Teacher Education,
Association of Colleges for Teacher Education
Division of Education Systems Development, United States Office of
Education
Professional Development Center, University of Houston
Teacher Corps

Copies of this report may be obtained for \$3.00 from

American Association of Colleges
for Teacher Education
Suite 610
1 Dupont Circle
Washington, D. C. 20036

National Dissemination Center for
Performance Based Education
150 Marshall Street
Syracuse University
Syracuse
New York 13210

This Conference was sponsored and this report was published as part of grants from the U. S. Office of Education and National Institute of Education, Department of Health, Education, and Welfare. The opinions and other content do not necessarily reflect the position or policy of the Agency, and no official endorsement should be inferred.

CONTENTS

COMPETENCY ASSESSMENT, RESEARCH, AND EVALUATION

FOREWORD	W Robert Houston, Conference Director and General Editor of the Report	
		Page
PART I	FRAMEWORK FOR RESEARCH AND DEVELOPMENT	
Chapter 1	Conceptual Model of R & D for CBE Frederick J. McDonald	11
Chapter 2	National Institute of Education and Research in C/PBE Garry L. McDaniels	22
Chapter 3	Closing the Knowledge Gap H Del Schalock	34
PART II	ASSESSING PROFESSIONAL COMPETENCIES	
Chapter 4	Listening to Different Drummers: Evaluating Alternative Instructional Models Bruce R Joyce	61
Chapter 5	Task Force on Relating Teacher Competence to Pupil Outcomes Donald M. Medley and David Krathwohl	82
Chapter 6	A Practical Approach to a Complex Problem George Gustafson and Marjorie Powell	88
Chapter 7	Classroom Observation Robert S Soar	100
Chapter 8	Inservice • Needs Assessment • Competency Based Teacher Education Gilbert F Shearron	112
Chapter 9	Assessing Teacher Competence Beatrice A. Ward, Greta Morine, and David Berliner	124

BEST COPY AVAILABLE

Chapter 10	Teacher Competency Research Barak Rosenshine	138
PART III	PROGRAM EFFECTIVENESS	
Chapter 11	of the Program Is Competency Based How Come the Evaluation Is Costing So Much?	155
	Michael Scriven	
Chapter 12	Task Force on Evaluating Program Effectiveness	164
	Rolace Aubertine and David Potter	
Chapter 13	Criteria for Describing and Assessing Competency Based Programs	168
	W. Robert Houston, Jr. Bruce Burke, Charles E. Johnson, and John H. Hansen	
PART IV	EFFECTIVENESS OF INSTRUCTIONAL STRATEGIES	
Chapter 14	Evaluating Ways To Help Teachers to Behave Desirably	173
	N. L. Gage	
Chapter 15	Task Force on Evaluating Instructional Strategies	186
	Winford A. Weber and Ned Flanders	
Chapter 16	Priorities and the Nature of Evidence	190
	Allen A. Schmieder and Jone Mark	
PART V	SOCIAL CONTEXT	
Chapter 17	The Politics of Teacher Competence	207
	Harvey B. Scribner and Leonard B. Stevens	
Chapter 18	The Decision-Making Process in Planning and Implementing Education Programs That Meet the Needs of Our Diverse Society	216
	Carl A. Grant	
PART VI	INSTITUTIONAL CHANGE	
Chapter 19	Some Unscholarly Views on Institutional Change	227
	Don Davies	
Chapter 20	Task Force on Institutional Change	232
	James Dyer and Dell Felder	
Chapter 21	Model for Developing a State System for Managing Performance Programs	234
	Robert C. Burkham	
Chapter 22	Implementation of CBTE Viewed as a Developmental Process	252
	Gene E. Hall	
AUTHORS		270

FOREWORD

The *sine qua non* of competency/performance based programs is assessment, research, and evaluation, yet in these areas remain the knottiest problems faced by educators today. They thrust educators beyond the cognitive into assessing the *performance* of teachers and other educational professionals. They press for measures of effectiveness in assessing the *consequences* of the professional's actions.

C/PBE provides the basis for developing assessment and evaluation techniques and employing them in research efforts. In a recent monograph, the Committee on Performance Based Education of AACTE noted that this movement "offers an opportunity for significant research because (1) increased precision in defining goals, identifying assumptions, and measuring outcomes greatly improves the conditions for analyzing cause and effect relationships within teacher preparation programs, and (2) the possibility of having groups of teachers with more clearly identified constellations of abilities enhances the probability that relationships between teacher preparation and pupil outcomes can be fruitfully studied."

Schalock, in proposing a research strategy as part of the efforts of the Consortium of CBE Centers, emphasized that, "At present we have only the barest of data on program cost when operating within the CBTE mode, and no evidence as yet upon cost/benefit relationships. We also have no firm evidence that one set of competencies is more productive of learning in children than another set, that one level of competency definition is any better than another level, that one approach to program organization and instruction is better than another in bringing about desired competencies, etc., etc. Again, questions of this nature can be answered only through research, and at this point in time such research has not and is not being undertaken." Albert Shanker, in a column in the *New York Times*, concluded that "Instead of hard data and controlled experimentation, educational panaceas are put forth on the basis of subjective impressions."

The importance of building this data base is acknowledged by all; within recent years, an increased effort can be perceived. Many C/PBE projects have integrated research into program development. Research and Development Centers are conducting series of studies, the National Institute of Education has planned a series of conferences and activities to devise a

BEST COPY AVAILABLE

thematic research strategy, membership in the American Education Research Association has increased rapidly, and other organizations are emphasizing assessment and research in their programs.

Certainly, these areas--assessment, research, and evaluation--are the most challenging in education today, and the most powerful catalysts to improved training of educational personnel. To consider the current state of the art, and to encourage further efforts, a conference was organized to study CBTE. The papers included herein were delivered at the National Conference on Competency Assessment Research and Evaluation held at the University of Houston on March 13, 14, 15, 1974.

The Conference was designed so that eminent educators and researchers could speculate on needed thrusts, consider current achievements, and interact on major issues. More than 250 invited participants attended.

The Report is organized into six parts. The first part reports three frames for conceptualizing research in CBE. Frederick McDonald presented the keynote address in which he outlined five major areas for research: assessing professional competence, program effectiveness; effectiveness of instructional strategies, social context; and institutional change. Each of these areas, in turn, was the focus of a major presentation, several clinics, and a task force. These presentations are included as chapters in the last five parts of the report, and are self-explanatory.

Thanks are due to so many persons who made that conference and this publication possible. The directors of projects and programs sponsoring the conference responded to numerous calls for assistance in identifying speakers, program personnel, participants, and in making fiscal and other arrangements. These sponsors are described in detail below.

The *Multi-State Consortium on Performance Based Teacher Education* is a federally funded project (Title V, ESEA; Teacher Corps, National Center for the Improvement of Educational Systems) designed to assist the participating states (Washington, Oregon, Utah, New Jersey, Arizona, Minnesota, Vermont, New York, Texas, Florida) in developing appropriate policies related to each state's unique approach to performance education. The Consortium also provides a national communication and dissemination resource through the publication of a monthly newsletter, *PBTE*, and a series of monographs. Theodore Andrews of the New York State Education Department (the administering state) is Director of the project.

The *National Commission on Performance Based Education* seeks to improve programs for the preparation of educational and other professional personnel by initiating, supporting, and coordinating PBE research and development activities. Direction for these activities will be provided through the implementation of a comprehensive Research and Development plan, with initial efforts focused on teacher preparation. The Commission is supported by a grant from the Rockefeller Brothers Foundation; Frederick J. McDonald, Educational Testing Service, is executive director.

The *National Consortium of CBE Centers* is an informal association of institutions involved in the development and implementation of Competency Based Teacher Education. At present, there are nine National CBE Centers: Florida State, Georgia, Houston, Michigan State, Syracuse, Teachers College-Columbia, Teaching Research/Oregon College of Education, To-

ledo, and Wisconsin. The Consortium serves as the coordinating body for the CBE Centers, serves as a conceptual forum focusing upon the refinement and advancement of CBE concepts, provides a source of CBE leadership at the national level, functions as a clearinghouse for providing developmental assistance to meet priority CBE needs, commissions papers and publications on unresolved issues, and sponsors think-tank symposia on key CBE topics. Norman Dodd, Florida State University, is the current chairman of the board of directors, and John H. Hansen is the executive secretary.

The *National Institute of Education* (NIE) was created in 1972 to support the policy of providing every person with equal opportunity to receive a high quality education. Under the leadership of Thomas K. Glennan, Jr., the Institute has established a number of priorities and programs to meet its goal.

The Program on Teaching and Curriculum is one of these major efforts. This Program's focus is to assist teachers in providing effective instruction to children. In addition to ongoing work in research on teaching and curriculum renewal, the Program is developing a targeted research and development agenda to guide future studies of teaching and curriculum.

The *Committee on Performance-Based Teacher Education* of the American Association of Colleges for Teacher Education gives direction to the Association's national project on PBTE. Composed of representatives of colleges and universities, state education departments and the organized teaching profession, it seeks to clarify the PBTE concept, study the state of the art, stimulate dialogue about PBTE's promise and limitations, provide practical help to operators of PBTE programs, and make recommendations which will assist in achieving the potential of PBTE. The major activities of the project include the publication of pertinent monographs and the sponsorship of various types of training opportunities. The project is supported by NCIES of the U.S. Office of Education through the Texas Education Agency, and is directed by Karl Massanari, AACTE.

The purpose of the *Professional Development Center*, University of Houston, is to facilitate the development of CBE programs for preparing professional personnel in Texas. The Resources Center in room 466 is part of the Center. In addition, it is cosponsoring with the Texas Education Agency and the Multi-State Consortium a series of regional conferences and provides support services for program development. The PDC is a special project of the Teacher Corps. W. Robert Houston is director of the PDC, Sarah White is associate director of the PDC.

Teacher Corps is dedicated to improving the education of children from low income families. It specifically includes authority to broaden programs of teacher preparation from prospective teachers, as well as those experienced teachers not in the schools. Now preparing for its ninth 2-year cycle, Teacher Corps supports some 80 projects in 35 states, Puerto Rico, and the District of Columbia. Approximately half of the projects are in major cities, and three-fourths are in elementary schools. The heart of each project is the Teacher Corps team of interns and experienced teachers working together with the schools, their communities, and their nearby universities in the planning and operating of innovative programs for the training of teach-

BEST COPY AVAILABLE

ers. The director of Teacher Corps is William Smith, and the associate director is Russell Woods.

Several meetings of the Coordinating Committee of the National Commission on PBE were devoted to this topic. Frederick McDonald, Theodore Andrews, Howard Coran, Allen Schmieder, David Potter, Horace Aubertine, Karl Massanari, and Robert Houston conceptualized the program design, suggested areas of research to be included, and recommended a number of the speakers for the conference.

Garry McDaniels and Frank Sobol identified topics for several sessions and encouraged many of the participants to attend.

All of the organizations listed above and particularly their directors, were influential in shaping the final program format, identifying educators who are concerned with research in CBE, and supporting the conference both financially and conceptually. James Steffensen continually provided ideas and a resilient backboard for ideas during the development stages.

Thanks too must be accorded to program personnel. Following the conference, many of the speakers reviewed and edited their presentations on the basis of questions and feedback. Task force synthesizers consolidated several hours of discussion into a few pages and several relevant questions. Their efforts helped make the conference successful, while their careful attention to detail has made this report more useful.

Dean Robert Howsam facilitated use of the College of Education at Houston, and provided leadership in conference activities. While more than 50 faculty, staff, and graduate assistants worked in the conference, the following deserve special identification as members of the arrangements committee and chairpersons of various committees: Eugene Chiappetta, registration; Barry Pelton, publicity; Joseph Schroeder and Jay Shores, audio-visual equipment and resources center; Ida Stewart, hospitality; and Bruce Thompson, transportation.

Obtaining manuscripts from authors, copy editing them, and preparing them for printing were accomplished by Barbara S. Coleman and Sheila Ford, University of Houston, and Anne Finch, New York State Education Department. Layout and format was designed by Pat Schrader and Parker Floyd of the University of Houston Printing Plant, and the publication was printed by the Multi-State Consortium.

June 1, 1974
Houston, Texas

W. Robert Houston
Conference Director and General Editor of the Report

FOOTNOTES

American Association of Colleges for Teacher Education. *Achieving the Potential of Performance Based Teacher Education Recommendations*. (Washington, D. C., February 1974).

Dr. Schaeck. *Closing the Knowledge Gap*. (A Position Paper of the Consortium of Elementary Model Directors). November, 1973.

BEST COPY AVAILABLE

PART I

FRAMEWORK FOR RESEARCH AND DEVELOPMENT

1 Conceptual Model of R & D for CBE

Frederick J. McDonald

There are two questions which are the foci for analysis and discussion at this conference given that there is a definable set of events called competency based education, what constitutes evidence that it produces its claimed effects and how much evidence is sufficient before we decide that it is an effective strategy for training teachers? When we bring together a group of people as diverse as this group, each brings his or her own ideas about what the significant research problems are in competency based education. What we are trying to stimulate as we go through this conference is, however, to focus your attention on those two questions.

In my opinion, the goal of our research and development efforts ought not to be to prove that competency based education is effective. That is an impossible task and can probably never be realized. However, if the problem is taken to be to establish what constitutes sufficient evidence so that competency based education may be established as a matter of social policy, then a research and development program is more readily conceived.

My role is to lay out what I think are some of the major problems and the major questions which set the network of problems on which evidence is needed. The logic of this presentation is to describe the problems for which evidence is needed. I will be answering the question, what are the critical questions which must be answered if we are to find whether or not competency based education makes a difference? In some cases, I will suggest the criterion against which the evidence ought to be projected. The question of how much evidence is sufficient is more difficult to answer, and I will consequently have considerably less to say about it.

The problem of what constitutes sufficient evidence that competency

based education ought to be the style and strategy for training teachers is part logical and part psychological. The logical aspect relates to the structure, amount, and kinds of evidence needed to constitute sufficient proof that the theory or model or strategy is useful. The psychological aspect relates to what individuals will demand because of their intuitive and preferential reactions to the concept. The psychological aspects of belief and acceptance are not totally responsive to the logic of data.

One component of the psychological aspects of stimulating conviction is that as competency based teacher education catches on, significant social implications for the teaching profession will be apparent. As a consequence, people whose lives will be affected by competency based education are looking at the political implications of certifying teachers on the basis of demonstrated competence. Their concern is twofold: first, that the concepts of competence will be defined independently of the conditions of teaching; second, that the criteria of competence will not take into account other factors that influence both teaching and learning.

The conclusion to be drawn from this concern is that research designs must include the variety of factors that probably impinge on teaching performances and their consequences in learning. It is reasonable to assume, for example, that variations in administrative structure, curriculum, and characteristics of children and teachers affect teaching performance itself and also learning. It is obvious that an important research problem in its own right is to measure the relative effects of such factors. Descriptions of competence divorced from the conditions of teaching and learning are not likely to be believable.

The criteria for assessing the sufficiency of evidence appear to be of two kinds. One set is related to the validity of the results of the research --what data support a hypothesis that a teaching performance is effective, and how generalizable are the results. The second set are more elusive, related more to the context and content of the research. The research must be believable in that it attends to the complexities and realities of teaching.

I address myself primarily to the kinds of knowledge needed to support the proposition that competency training is the most effective way to educate teachers. I will discuss several basic research problems whose solution is required if the effectiveness of competency training is to be demonstrated and accepted. Other speakers will discuss these problems in greater detail, elaborating their conceptual and methodological difficulties.

Let me begin with what I think is the basic problem, which happens also to be fraught with implications of the kind previously described: that is, what is the nature of teaching competence? Some of us seem to assume that because we have more precise descriptions of teaching competencies that we have resolved the question of what is teaching competence. One of the most commonly stated accusations aimed at competency education is that we believe that a list of behavioral descriptions of competence constitutes a description of competence. We are quite properly asked, "How do you know that if the teacher performs as described that the performance will have a desirable effect on student learning?"

A competence may be described as a behavior of a teacher, for example, asks students to give reasons for their opinions on controversial topics

being discussed in class. Presumably, a high frequency of such statements indicates that the teacher has the behavioral performance in his or her repertoire. But is such information evidence of competence? It is only if we assume that the appearance of the behavioral act is synonymous with competence.

Such an approach is rightly charged, in my judgment, with indifference to the consequences of the event. The more cynical of critics point out that this concept of competence leads logically to certifying as competent teachers who emit the behavioral performances in the absence of students. While such criticisms are unnecessarily harsh, they point to the problem: is competence to be defined in terms of effects manifested in students, in terms of observable acts of teachers, or in terms of pairs of acts and their effects?

It seems to me that we can no longer avoid saying that teaching competence must be defined in terms of effects on pupil learning. Unless the research and development movement behind competency based education addresses the problems implicit in that concept of competence, it seems to me there is no hope of ever persuading anybody that the competency based education movement is doing anything other than developing semantic descriptions of what constitutes competence. I propose that the first and most important research problem to attack is the question of what teaching performances produce significant effects on learning.

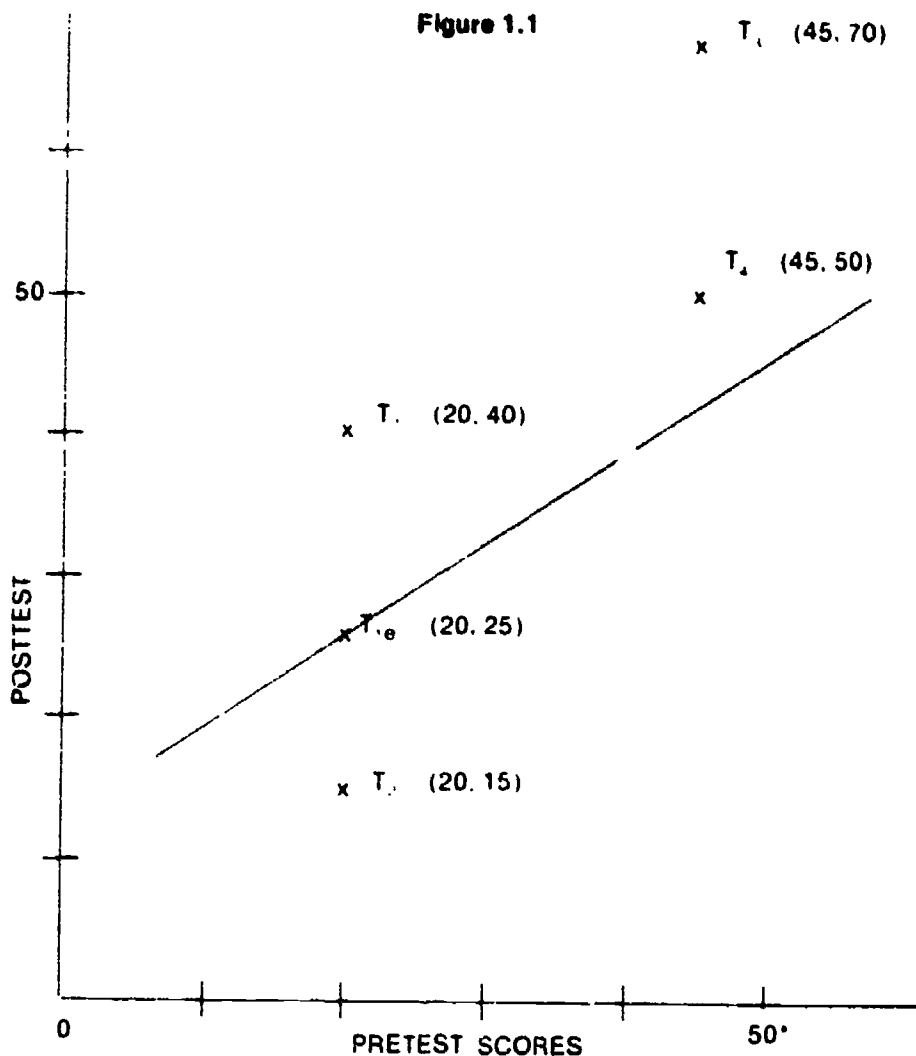
There are two general research methodologies that may be used for this purpose. I would like to describe them very briefly. The first of these is an inductive approach to the analysis of the relationship between teaching performance and student learning. This methodology proceeds as follows. The research should be conducted so that actual teaching may be observed, preferably that of more experienced teachers. You need not have preconceptions about what kinds of performances are likely to affect student learning, or your observations of performance may be guided by theory or previous research. But you must have a clear and specific conception of the kinds of student learning that you are going to study.

The simple form of the methodology consists of measuring student learning at two points in time.* Given two such measurements obtained on each of the students in the classes of a sample of teachers, a regression of the later or posttest scores is performed on the earlier or pretest scores. The results of this statistical analysis can be diagramed as in Figure 1.1.

The points are represented by the mean scores for a teacher's class, the first number being the mean for the pretest, the second, the mean for the posttest. The line through the points is called a regression line and points on it represent expected scores. Given the information in the analysis, we expect T_i 's posttest mean score to fall on the line at T_{ie} ; similarly, we would expect T_j 's posttest mean to be located on the line at the same point. Both posttest means, however, locate T_i above the line and T_j below the line.

* Learning in this speech is an omnibus category meant to include all desirable changes in students' cognitive, affective, and social. Nor am I concerned about the method of measurement or the difficulties of measuring some changes. The researcher needs to be specific about the changes desired and have developed reliable and valid ways of measuring them.

Figure 1.1



Regression Analysis of Posttest Scores on Pretest Scores Using the Scores of the Students of a Sample of Teachers

*The metric is arbitrary. The first number in parentheses represents the mean score for the T's class on the pretest, the second number, the mean on the posttest.

When you do this type of analysis, you are looking for teachers who depart markedly from the regression line because you want to study them intensively to find how they differ from teachers who are closer to the regression line. The general idea is that teachers whose student performance shows marked departures from the line may be highly effective, other things being equal. We ignore here the many technicalities of this type of analysis to make the point that the method provides a way of identifying teachers who may be unusual and provides a way of inducing hypotheses about what constitutes performance related to student learning.

The next step is to study the observed performances of these teachers and find the performances which discriminate their teaching performances from that of other teachers in the analysis. If you gather enough information around the performance of the children and the performance of the teacher, such as the kinds of schools in which they are, the composition of the class, and similar factors, you stand a fairly good chance of sorting out those performances associated with the act of teaching which distinguish more effective teachers from other teachers.

It is not true, however, that these performances are guaranteed to be the ones that are associated with effective teaching. Let us consider the problems of this type of a study with respect to the two questions that I asked earlier: what constitutes evidence, and how much evidence is sufficient? At the end of this type of study, all you have is a set of hypotheses about the relationship between teaching performance and student learning that may have some validity. These hypotheses must be tested either by doing experimental studies or by doing replications of the original study or both.

The problem of how much evidence is sufficient arises from the complexity of these studies. It is feasible to study a limited number of student outcomes. To gather data on students on relatively few outcomes requires several days of testing. To study the teachers over time requires multiple observations. The question is, how many studies of this kind are needed before we have sufficient evidence of the components of competence? Where are they to be located in a matrix of pupil changes, types of schools, differences in organizational structures, differences in levels of education, and differences in teacher and student characteristics before we have confidence that we have working hypotheses that describe competence?

What we must not do is propose or claim that we need 500 of these studies and that it will take 10 years and \$10 billion of research money to describe teaching competence. We must recognize that we are living in an environment of pessimism about the probability of demonstrating that there is a relationship between teaching competence and pupil learning. One of the goals of the research and development program on competency based training should be to demonstrate the *probability* that teaching competence can be identified. It seems to me that several studies of the kind that I have described are likely to demonstrate some relationships. Such studies will at least eliminate some hypotheses about teaching or indicate weak relationships. Because such studies involve correlational designs, however, and because they are inductive in character, potentially valid hypotheses may not be discovered. Another type of strategy is needed that is theoretical-experimental in character.

A theoretical-experimental type of study varies teaching performances systematically and assesses the effects of these variations on student performance. The manipulation of the performances may be of two kinds. One manipulation is to vary the degree of the skill of a performance to determine the relation of the degree to the amount of the effect. Such experiments provide the data necessary to establish a criterion level for the performance.

A second type of experimental manipulation requires varying the kinds of performances to assess their effects on the same outcomes. Such studies

provide comparative data on the effects of performances, their interchangeability, and their interdependence

These experiments should be conducted under controlled conditions where the conditions and materials for teaching are identical and where students are assigned randomly to classes. They may also be conducted under field conditions using fractional factorial designs. This type of design involves selecting relevant contrasting conditions in which the teaching occurs such as those associated with the characteristics of students, schools, or curricula. Fractional factorial designs are notably efficient because relatively few teachers can be used to test a relatively large number of hypotheses. When used under field conditions, they provide tests of interactions among school and pupil factors and teaching performances.

The advantage of experimental designs is that they provide a way of testing hypotheses systematically. Experiments are powerful ways of testing theories of teaching by testing the hypotheses derived from them. One can begin as follows: select one or more models of teaching; define or describe the kinds of pupil changes that the theory predicts will occur if children were instructed by teachers applying that model; and derive hypotheses about what kinds of teaching competencies produce these changes in students. The experiments are then designed to test these hypotheses.

The advantage of using theories or models as starting points for deriving hypotheses is that the size of the hypothesis-pool will be greatly enlarged. The disadvantage is many hypotheses will be derived of unknown validity either because they are freshly derived from the theory and have no empirical support from any source or because the theories themselves have weak empirical support. At this stage of research on teaching, it is probably better to include more rather than less hypotheses in the pool for testing.

A research and development strategy on teaching competence ought to include both inductive-correlational and theoretical-experimental studies. Further, each inductive-correlational study ought to be followed by experiments which test the validity of the hypotheses derived inductively.

The problem remains, what constitutes sufficient evidence to make reasonably definitive statements about teaching competence? One approach to answering this question would be to consider the consequences of training for a set of competencies that has been derived from an inductive-correlational study. Here the evidence is relatively weak—no replication of the original study and no experimental verification. There seems to be no risk in using such information, particularly if the training program studies what happens when its teachers use the competencies in real teaching situations. In this case, the opportunity exists to generate more data about the effects of the teaching performance.

The problem of sufficient evidence seems to lie in the political arena. If a state agency *requires* that programs train for the competencies derived from even a few inductive-correlational studies which have not been replicated, one has cause for concern.

It also seems unwise to assume that there will be an end-point at which it can be said that we now have sufficient evidence to "justify" competency training. Rather the hypothesis-generating, hypothesis-testing life style should characterize the development of the competency based movement.

BEST COPY AVAILABLE

There ought not to be a time when we have stopped research on teaching competence.

A few general comments about the problem of studying teaching competence are relevant here. In studying teaching competence, we have to separate out systems' effects on children from teacher effects. Many of the goals of education are broad. Changes with respect to them occur as the consequence of the teaching of the different kinds of teachers whom the child encounters and also as a consequence of other kinds of influences, such as those from peers. It seems to me that our basic problem in generating a hypothesis-pool about teaching competence is to generate it in such a way that the systems' effects are treated as potential moderators of teaching performances. Research on competence ought not to be, at least initially, research on effects that occur as the consequence of five years of education. Such research is both interesting and important, but it is not really related to demonstrating the effects of teaching competence. When we know more about the specific effects of teaching acts under known conditions, we will be prepared to study systems' effects.

It is important to keep in mind that in conducting the type of inductive field-oriented research that I referred to earlier, it is possible to design those studies so that a wide variety of system effects may be studied as interacting with the effects of teaching performances. You can determine the kinds of influences that the administrative climate of the school, for example, has on teaching performance, and how performance varies within different organizational structures or within different types of programs. But our problem is to account for, or to describe how much of the variation in student performances actually can be accounted for by various teaching performances. The goal of our research is to find the effects of teaching; other factors in this research are important if they influence or so affect learning that the influence of teaching is minimal. If we cannot determine the effects of teaching performances, the competency based movement is irrelevant.

Another problem is that we need to know how different kinds of teaching performances cluster ---we know practically nothing at all about the structure of teaching performances. There is some evidence that specific descriptors of performance cluster in factors. Most descriptions of competence are lists of behavioral acts. Interdependencies among them are made logically. But if such specific acts tend to occur together in groups, and if these clusters have demonstrable relations to changes in students, then these clusters are probably the basic units of teaching performances.

Then the question is, what is competence? Is it the performance embraced in such factors, or is it the basic descriptors that are spelled out specifically and that an observer may check as present or absent in a teaching episode? We really do not know, but some evidence suggests that perhaps it is these clusters and not the most miniscule aspects of the total performance that produce learning effects. The inductive-correlational type of study provides the empirical data required for the analysis needed to study this problem.

A fundamental problem facing us from a research point of view is: what is the unit of competence? Is it the simple semantic descriptor which is

BEST COPY AVAILABLE

useful in devising a training program or is it some combination of these units that produces the expected effects of a teaching performance?

Permit me to summarize very briefly. The focus of this presentation has been on the problems of research on teaching competence, which are problems of designing a strategy to identify and isolate the effects of teaching performances. This strategy should include two kinds of studies, inductive-correlational and theoretical-experimental. In all such studies, the criterion of effective teaching must be changes in students. Second, a judicious combination of designs will facilitate testing a very large number of hypotheses in a relatively short period of time.

Inevitably, any conference discussing research and development programs and problems in teaching, particularly a program directed to creating an empirical base for competency based training must face three facts. First, there is a body of research whose results are mediocre and meager. Second, there is a body of literature and a school of thought that seems to support, in the case of the literature, and has concluded, in the case of the school of thought that schools and teachers do not make a significant difference in learning. Third, there is widespread pessimism that research on teaching and specifically on competence in teaching is ever likely to produce useful knowledge.

The import of my presentation is that irrespective of what is taken as the present state of the art in research on teaching, methodologies may be systematically applied which are likely to identify the parameters of competence. It seems pointless to continue to bemoan correlational studies when it is obvious that their logical weakness in establishing causality can be ameliorated by replications and experimental studies. Also, the usual jeremiad that partialling out the influence of school and other factors is difficult seems pointless when so few researchers study teaching and its context. Lest I slip into a moralistic diatribe, suffice it to say that our problem is to design a strategy that attacks the study of teaching competence systematically and with confidence that meaningful results are likely even if they may not be overwhelming.

There is one problem in designing such a strategy that will not yield to rational analysis alone. The research problem, as I have described it, requires first a decision about what pupil effects will be used as the criterion. Occasionally this choice will be predetermined as when a political agency stipulates pupil outcomes for which they wish to determine the antecedent teaching performances. Otherwise the choice of outcome is made by the researcher. Thus, we must as a research and development community arrive at some consensus on the priorities of these outcomes.

I have chosen to place the description of competence as the critical problem on which research must be done. It should not be inferred that all other activities need be postponed until considerable progress has been made on this problem. Competency based training is here both in the form of program development and state mandates requiring shifts to competency based training. Two other kinds of problems require research.

The first of these problems is the effectiveness of training modules. That research is a straightforward engineering problem. The research must be designed to show that a particular module has produced a change in train-

ees which it is expected to produce. One component of this research is specifying the conditions under which you will decide that the trainees should demonstrate that they have acquired a skill. For example, if you use microteaching you may demonstrate that under the conditions of microteaching trainees have acquired a particular performance. Is that sufficient evidence for deciding that the trainee has acquired a skill? Many of us are accepting the criterion that at the end of the module a prespecified performance must be demonstrated by some assessment procedure. That criterion is sufficient for deciding that the experimental treatment designed into the module does produce the desired outcome. We do not know, however, whether or not the trainee can manifest the performance under other conditions. At some point, we must demonstrate the links between these specific acquisitions and the ultimate criterion of effectiveness—classroom performance. Hence two kinds of research are needed: 1) does the module produce the specified change; 2) do performances acquired at the modular level appear in complex performances in realistic teaching situations?

Assume that a module does produce an immediate change but appears to have little influence on subsequent classroom performance. One conclusion is that the module has no transfer value to teaching. But the problem may be within the system of modules. Many are assuming that the acquisition of teaching skills is an additive process—that components are added from the simplest to the most complex. This view is useful for an initial organization of a system of modules.

The modules may facilitate transfer from the acquisition of one skill to another or may inhibit such transfer. Some arrangements of modules may be more efficient than others. Some skills may be better acquired in their complex form rather than by accruing components of the skills. Thus, once one designs an instructional system, its assumptions about the composition of complex teaching skills should be explicated as well as the assumptions being made about how the skills may be best learned as the trainee proceeds through the system.

Each system is a model of the processes by which teaching competence is acquired. These models ought to be compared by treating them as experimental conditions.

Thus, a major research area is the problem of instructional effectiveness. We need to know not only what constitutes teaching competence but also the most effective ways to educate for it. Research in this area should ultimately answer two fundamental questions: what aspects of teaching competence are learnable and under what conditions can they be acquired most efficiently?

A related problem is the evaluation of program effects. Do competency based programs produce better teachers? Attempts to conduct research on this problem at this time are premature for three reasons. The most obvious is that there are not sufficient graduates of such programs to yield worthwhile evaluation data. Further, until programs are better developed, there is little point in conducting comprehensive evaluations of programs. Second, the criteria for evaluating program effects are neither clearly defined nor agreed upon.

Third, if competency based programs are to be evaluated, the design of

such evaluation ought to provide for comparisons of planned variations in the training programs. As the preceding analysis hopefully suggests, there is ample room for varying conceptions of teaching competence and for systematic variations in program design. We ought to anticipate program evaluation by planning for such designs now. These problems must be solved within the next few years, consequently, problems of program evaluation is a major topic of this conference.

This discussion of research and development in competency based education has focused on two major problems, and by implication on a third---the problem of measuring teaching competence. Teaching competence cannot and is unlikely to be measured in any meaningful way at the present time because we do not know what it is; however, research on the nature of teaching competence and the effectiveness of instructional systems should hasten the development of the technology of assessment. Many people believe that finding ways of measuring competence is urgently needed. But, in my opinion, they seem to be seeking panaceas. There are no tests, observation systems, rating scales, or even simulated teaching situations that can be used in any sensible way except in the context of research on competence. Thus, a major concern of this conference is the use of assessment procedures in research on competence and instructional and program effectiveness.

This conference will also give attention to two other problems that require study. The first of these is related to the social context of the schools. Teaching skill is influenced by the composition of the staff and students of a school, by the organization of the staff and students of a school; by the social milieu of the schools, by the value consensus of staff, students, and community, and by curriculum designs. Thus, the study of competence is necessarily incomplete until it is studied in the context of schools, traditional or alternative, open or structured.

Teaching competence has multiple social meanings, and these meanings define the parameters of research on teaching effectiveness. Both the criteria and conditions of effectiveness vary as the social context of education varies.

The last problem that this conference will consider is the public policy, or more mundanely, the political implications of competency based education. There are two categories of problems to be considered in this respect.

First are the problems of change within training institutions. Competency based programs require changes in roles and habits, the structure of interpersonal relations among faculty and students, and the assumption of public accountability not heretofore either expected of or required by the training institution. We need to know the factors that facilitate or interfere with such changes, the problems encountered in change, and the most effective ways of resolving them.

Second are the problems of public policy. These problems relate to the control of the training of and certification of teachers and the criteria for and evaluation of teacher effectiveness. The mandating of certification on the basis of competence is a matter of public policy with significant implications for institutions and organizations whose interests are affected by such major shifts in public policy. We know very little of a systematic char-

acter about these political problems, though some of us may have practical experience with them. It is difficult at this time to visualize what research needs to be done, but the competency based movement is immersed in the political process. What issues of this character need to be studied and how should they be studied?

In conclusion, this conference will address these five problems: 1) what is teaching competence, 2) what are the most effective ways of educating for it, 3) are competency based programs producing effective teachers, 4) what are the influences of the social context on teaching competence, and 5) what are the public policy issues relevant to competency based education?

The purpose of the conference is not to provide answers to these questions because there are no simple answers to them. Its purpose is to clarify the problems, to propose appropriate research strategies for studying them, to bring current experience into the analysis of the problems, and to stimulate interest in a research and development program that will undergird the competency based movement.

Today is hopefully the beginning of the creation of a national research and development effort whose ultimate purpose is through the study of teaching and its effects to influence the quality of education and to provide the knowledge necessary to make the most profound and far-reaching processes for change in teaching successful.

2 National Institute of Education and Research in C/PBE

Garry L. McDaniel

The purpose of this conference is to focus on the question, "What are the things that teachers do which make a difference for children?" This question is reasonable. Our business is to help children; however, as unreasonable as it may seem, the quantity of research linking teacher training and teacher performance with child performance is limited; some would say, nonexistent. The National Institute of Education (NIE) plans to ask this question repeatedly over the next several years in its Program on Teaching and Curriculum. In asking this question, we hope to examine many of the assumptions underlying competency based teacher education and licensing.

In the pages that follow, the process by which the NIE will plan its research efforts in this area will be described. The first section argues that the current relationship between funding agencies and researchers could be made more efficient. The second section describes how to establish a more efficient relationship between researchers and funding sources. The third section describes one phase of NIE's planning efforts—the National Conference on the Research on Teaching.

While these efforts are more broadly focused than the issues related to competency based licensing and certification, parts of this effort will obviously be relevant to the primary concerns of this National Conference on the Assessment of Competency Based Teacher Education and Licensing.

The Rationale for Development of a Research Plan

The purpose of this section is to describe the efforts of the National Institute of Education to create a relationship between itself as a funding agent and the research and development community. While the focus is on

NIE and its relationship, the principles apply to the more general issue of the relationship between any funding agent and its constituency

The assumption underlying this discussion is that relationships between funding sources and researchers, especially in the social sciences, are ready to become more efficient, that these more efficient relationships will improve the impact of social science research on contemporary problems, and that improved relationships will also enhance the image of the social sciences as an effective tool for attacking social problems. It is NIE's premise in this planning conference that a major path toward improving efficiency will result from improving the ability of researchers and funding sources to develop research plans

The Current Relationship Between Most Funding Agencies and Their Related Research Communities

At present, the process for funding education research is as follows. The researcher develops proposals in isolation from the funding source. Proposals are sent to several funding sources where some rather abstract objectives seem consistent with the intent of the investigator. If a source finds something interesting in the proposal, a peer panel is usually assigned (a) to assess its internal quality, and (b) to consider the merits of the proposal as it relates to some "field." Finally, a set of the most highly rated projects are matched against a budget, a budget usually determined prior to the submission of ideas.

Several sources of inefficiency can be seen here. The first is the problem of not knowing where to send a proposal. The second is that, even if the appropriate funding source is found, it is necessary not only that the general content of a proposal be adequate, but that the proposal emphasis be related to the primary concerns of the funding source. Often, the proposal is wide of the mark at the time of the evaluation. The third source of inefficiency is that a collection of good proposals that add up to a program strategy must often be arbitrarily cut because a budgeting level was set prior to the establishment of strategies for attaining the goals of the funding source.

The efficiency and effectiveness can be improved if four essential types of activities occur

- The abstract goals and general priorities of the funding agent should be established
- Specific objectives and strategies should be developed
- Individual projects being funded should meet criteria of internal quality as well as criteria which require that one project relate to others in a cumulative fashion.
- All projects, strategies, and objectives should be disseminated in public documents so that public recommendations and budget allocations come before the actual funding.

It will be productive to consider the parameters of these activities.

Establishing Goals and Priorities

How carefully are the objectives of a funding agency stated? Let us consider, for example, the legislative mandate of an organization such as the National Institute of Education. It is broadly stated. The authorizing legislation asks the NIE to

- Help solve or alleviate the problems of, and achieve the objectives of, American education
- Advance the practice of education as an art, science, and profession
- Strengthen the scientific and technological foundations of education
- Build an effective education research and development system

Many education research sources, especially private sources, do not often define goals beyond the abstract level of a mandate. The Institute has recognized this problem. Therefore, one of the initial major actions of the National Council on Education Research (NIE's policy-making body) was to increase the specificity of NIE's mandate by identifying priority areas.

The priorities focus the work of the Institute in five areas:

- **Essential Skills** - investigating ways to aid all children in obtaining skills essential to function in society. Initial emphasis will be placed on reading. (\$1.1 million)
- **Productivity** - ensuring the Nation's continued ability to meet its commitment to quality education at a price it can afford, through improvements in the productivity and efficiency of the education system. (\$16.5 million)
- **Education and Work** - improving our understanding of the relationship between education and work, and improving career access and progression. (\$21.3 million)
- **Problem Solving** - accumulating the knowledge and experience necessary to assist state education agencies, communities, and their local schools to build the capacity to address local problems and adapt solutions suited to local needs. (\$8.6 million)
- **Diversity** - improving our understanding of individual and group differences in learning styles, needs, and preferences and developing programs that take those differences into account. (\$20.3 million)

While these priorities reflect progress toward specificity, the statements describe only the general character of NIE's concerns; they do not articulate the essential character of the problems being attacked by the funding source.

Developing Strategies

Funding agencies seldom articulate strategies for achieving their objectives or provide a rationale for the strategy adopted. It is exceptional to find a strategy clearly articulated in advance of making funding decisions. This point can be illustrated in some of the major curriculum development efforts. These efforts have led to the question of what to do next. There is some skepticism about the extent to which the national curriculum efforts

such as new math provided children with new mathematic understanding and skills

Assuming that the skepticism is reasonable, researchers have three general strategies regarding what to do next. One strategy is to do nothing if the outcomes are uncertain: let us no longer support large curriculum efforts, let us focus our efforts, for example, on helping teachers to develop their own curriculum. A second general strategy is based on the argument that such projects may not have really been well implemented in classrooms. Teacher training may have been inadequate, or materials may not have been sufficiently prepared. The issue becomes the process of implementation. A third strategy is based on the arguments that we really do not have measurement devices which are sensitive to such a sophisticated curriculum variation. This problem is essentially one of measurement. This third strategy argues that measurement issues must be given high priority.

While this is not an elegant illustration, it indicates how a public airing of one or more strategies would be helpful in getting the field of researchers to debate a common issue. And when a strategy is selected, knowledge of that strategy would improve the efficiency of the relationship between researchers and the funding source.

Evaluating and Interrelating Research Proposals

The third activity that should occur in developing plans is the development of strategies for evaluating research projects and proposals as to their (a) internal quality or logic, and (b) the extent to which one project relates fruitfully to other projects.

The first criterion has had primacy. The quality of the internal logic, design, analytic strategy, etc. provides the essential basics for project evaluation. But there should be a second criterion.

The second criterion receives notice but rarely any systematic analysis. Any evaluation of the degree to which projects are cumulative is usually left to those who review the literature, and such reviews are properly suspect. It is even difficult to organize symposia that put together projects that are additive or that produce a substantial mass of information related to a topic. Without a cross-project analysis, prior to funding, a cumulative impact is left to chance. There is the continuing assumption that if a project is intelligently designed and is funded, that project will somehow complement the next project funded. Even more optimistically, it is often assumed that, when all the projects are completed, the magnitude of some major problem will be reduced.

Communicating Projects, Strategies, and Objectives

The fourth and last activity needed is communication to make certain that a research plan exists which can define the relationship between funding agencies and their research communities. Sufficient and timely information regarding the objectives, strategies, and evaluation procedures of a research agenda is seldom available. There have been few, if any, instances in social science research where a public debate about the essential elements of a research agenda has occurred prior to funding. Although a

catalog of funded projects is usually available, it is rarely possible to obtain a catalog of proposed projects

Establishment of an Alternative Relationship

It is possible to create a relationship between funding agencies and their related research communities based on the foregoing four kinds of initiatives. An approach is being tested in the Program on Teaching and Curriculum (directed by Garry L. McDaniels) and in the Essential Skills Program (directed by Marshall Smith) at the National Institute of Education. After the general approach is described, the specific initiatives being taken will be presented. The general approach has four elements: (a) specifying major problems, (b) obtaining simultaneous input from field researchers regarding possible projects, (c) establishing alternative strategies for attaining goals, and (d) communicating alternative strategies with project recommendations.

Specifying Major Problems

Instead of having the funding agency identify major problems at a broad or abstract level, the agency must force concrete definition of problems. A Federal funding agency cannot remain at the general level of the goals of the authorizing legislation, just as the private foundation cannot effectively use the broad goals identified in its request to the Internal Revenue Service for tax-exempt status. Also, if the character of the problems are not fully described, it is also rather difficult to assess the accomplishments of the effort.

Obtaining Simultaneous Input from Field Researchers

Investigators are storehouses of project ideas. As time and conditions allow, an idea is selected, refined, and sent to a funding source by an investigator. Three problems result:

- The ideas are sent from different people at different times
- It is difficult to relate projects across agencies and foundations
- Each person's proposal is developed in a time lag from the recent work of other investigators; therefore, he may plow the same ground. If cross-project analysis is to be possible, it is essential that simultaneous input from field researchers be achieved.

Establishing Alternative Strategies

Not only must the input from field researchers be obtained simultaneously, it must also be obtained in a manner which identifies the intellectual and administrative relationships among projects. The development of a plan requires that the potential research ideas be organized into a logical agenda. Further, the organization should demonstrate how the projects relate to the achievement of an objective worthy of public or other support.

Communicating with the Research and Development Public

Finally, the steps described above provide information to be communi-

cated to the research and development public. Much of that public will be informed as a result of having been a part of the process. The remainder will have the time required for receiving and analyzing the pertinent documents.

The National Conference on Studies in Teaching

To further clarify the implications of these elements, the steps being taken in the Program on Teaching and Curriculum will be described. In regard to specification of major problems, the staff is meeting with students, teachers, administrators, teachers in colleges of education, etc. to get a summary of problem statements. These problem statements will serve as a charge to the researchers who will make project recommendations during the planning conference. Perhaps it would be ideal to have everyone meet simultaneously. Such a meeting would, however, require a quota system for representation of the relevant groups—which would result in huge working sessions that would be impossible to manage.

To obtain simultaneous input from field researchers, the present conference has been planned. To establish the organizational framework for the development of the plan, the panel chairmen will spend several weeks in making a separate analysis of subject area objectives, in cooperation with the conference participants. During the conference, the participants will refine those approaches, identify possible areas of research to implement the approaches, and specify the research projects which should be considered by the Agency.

Finally, to communicate these proceedings, several steps will be taken. First, all participants will leave the conference with complete volumes of panel suggestions. These will be edited for minor errors and sent to the Government Printing Office for printing and distribution. The analyses by the NIE which result in strategy recommendations will be available for general distribution at the end of August 1974. The logistics of printing and distribution have been established.

Substantive Form of the Panels

Preliminary discussion by NIE staff and consultants since August 1973 has fielded a formulation of the substantive structure of the conference—and the consequent foci of its 10 panels. In part, this structure is based on a categorization of major concerns in the career of the teacher: (a) recruitment, selection, and retention; (b) training, performance, and effectiveness; and (c) utilization. In addition, three foremost concerns of all Research and Development in teaching will be taken into account by panels on Personnel Roles in New Instructional Systems, Research Methodology, and Theory Development. Thus, the Conference will consist of the following panels:

1. Panel on Teacher Recruitment, Selection, and Retention
- 2-6. Panels on Teacher Training, Performance, and Effectiveness
 2. Panel on Teaching as Human Interaction
 3. Panel on Behavior Analysis in Education
 4. Panel on Teaching as Skill Performance

BEST COPY AVAILABLE

- 5 Panel on Teaching as a Logical and Linguistic Process
- 6 Panel on Teaching as Clinical Information Processing
- 7 Panel on Teacher Utilization
- 8 Panel on Personnel Roles in New Educational Systems
- 9 Panel on Research Methodology

In the following paragraphs, each of these panels is briefly characterized.

Panel 1 Teacher Recruitment, Selection and Retention

The panel on Teacher Recruitment, Selection and Retention will be concerned with planning Research and Development on devising, validating, and installing methods for improving formal and informal decisions at various points in a teacher's career.

- (a) admission to a teacher education program
- (b) retention in the program
- (c) admission to student teaching or internship
- (d) graduation from the program
- (e) certification or licensure
- (f) employment by a school system
- (g) granting tenure
- (h) promotion in rank and pay

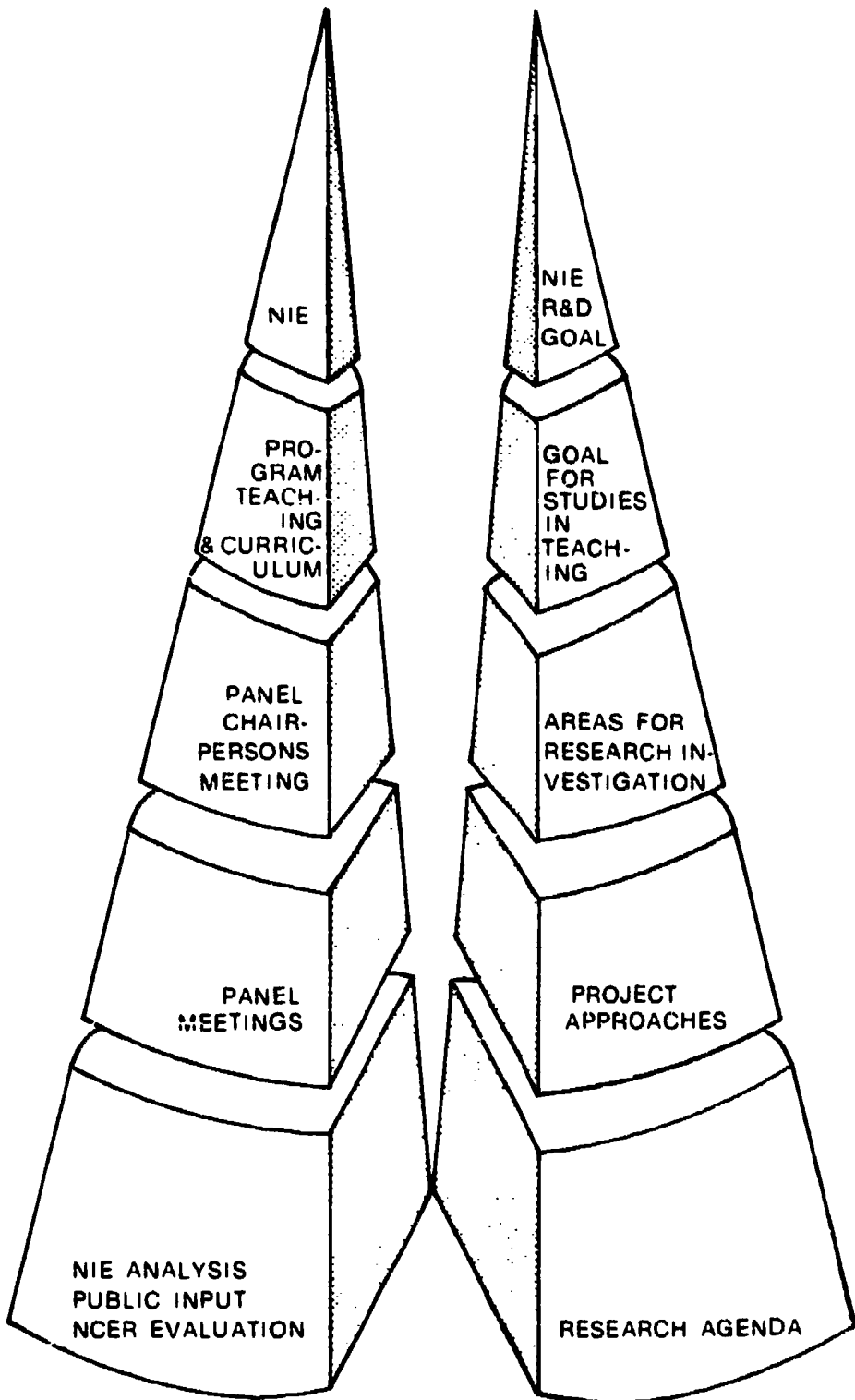
The classical paradigm in Research and Development on recruitment, selection, and retention consists of identifying criteria, selecting potential predictors of those criteria, and determining empirically their predictive validity. The panel will accordingly pay much attention to the matters of criteria, predictors, and ways of basing decisions on their relationships. The criterion problem will entail consideration of competency assessment and teacher evaluation procedures.

The panel will also deal with self-selection—ways in which persons are attracted toward or turned away from teaching, i.e., the occupational choice and attrition processes. It will examine the social class, ethnic, and other origins of teachers and the differences these make. It will also deal with the possibility that success in a training program is more a function of the program than of the candidate and his characteristics.

Panel 2: Teaching as Human Interaction

Many studies have dealt with observed human interaction (either live or recorded) as independent and dependent variables. In some cases, observations of human interaction have been made merely to describe differences between kinds of teachers in various settings and contexts. These studies have been reviewed in several monographs. This approach has been the most frequently used in research on teaching during the past two decades.

To review the current state of this approach, to lay out the most promising next steps (such as projects on taxonomies of teacher behavior, chains and patterns of teacher and student moves, computerized storage and analysis, and archival banks of videotaped interaction records), and to formulate



those projects in optimally definite terms—these illustrate the tasks that can be undertaken by a panel in this highly cultivated, but far from exhausted, field.

* Research and development on teacher training, performance, and effectiveness deals with teacher variables (teacher characteristics and behaviors) in either or both of two ways. (a) as dependent variables (with teacher training methods serving as independent variables), and

(b) as independent variables (with effects on students serving as dependent variables)

Work in this area has been based on a number of paradigms, each relatively independent. Hence, several corresponding panels (2-6) will work here

Training is construed to embrace socialization processes that imbue the trainee with the values and norms of the profession. It includes both preservice and inservice training. These panels will emphasize aspects of training that require practice, i.e., training intended to provide "knowledge how," not merely "knowledge that." The panels will be encouraged to develop new paradigms, cutting across or departing radically from present ones

Panel 3: Behavior Analysis in Education

Somewhat more recently, but with equal vigor and productivity, the behavior analysis in education approach has been applied to the improvement of teacher training, performance, and effectiveness.

The success of teacher training and especially of classroom management by these methods has been frequently reported. Thus, this thriving line of attack on several important problems in teaching and teacher training merits the attention of a panel. This panel might seek to increase the attention already being given to private events and cognitive processes. It could advance the effort to apply behavior analysis in education to the achievement of humanistic educational objectives. It could consider studying the effect of students on teachers through the cues provided by students, perceived by teachers, and responded to by teachers, and thus the ways in which students may reinforce teachers' responses.

Questions have been raised about the morality of "controlling" behavior through behavior modification methods. Hence, this panel may undertake to address such ethical issues.

Panel 4: Teaching as Skill Performance

During the past decade, an analytic approach to the training and evaluation of teachers has frequently been adopted. In such an approach, teaching is broken down into component skills, such as questioning, explaining, listening, and reinforcing. Little consensus has been reached on what skills should be identified. But the approach itself has flourished, lying at the basis of such innovations as microteaching and minicourses, in which teachers are trained in only one or a few skills at a time.

The analytic approach has also led to the movement to develop teacher-training products. Such products call upon the trainee to *practice* a given skill, in addition to reading, listening, or watching material about the skill. Hundreds of such products have been developed and cataloged, in part as a response to the nationwide attempt to make teacher training and evaluation "performance based." Their promise will be realized, however, only to the degree that they benefit from further research and development. The analyses should become more valid, logically and psychologically.

Panel 5: Teaching as a Logical and Linguistic Process

Studies of linguistic and logical aspects of classroom interaction here and in England have varied widely in the particular aspects of language studied, the kinds of classrooms selected for research, and the methods used for data collection and analysis.

Aspects of language studied have included the logical operations engaged in by teachers and pupils; the types of pedagogical "moves" in the classroom interaction "game" and their distribution among teachers and pupils; formal speech registers used in the school, as contrasted with the home, and their various special characteristics in various school subjects; and the sociolinguistic rules operating in the classroom seen as a special setting for face-to-face interaction.

Most research has been conducted in teacher-structured recitations, less in small-group and pupil-led discussions and group projects. Many researchers have assumed a set of norms governing classroom interaction that is shared by all participants; others have deliberately worked in multi-ethnic settings and considered variations in these norms as a special problem for analysis.

On the basis of this previous research activity, this panel will undertake to formulate the direction in which research should move to provide improved understanding of logical and linguistic phenomena in classroom settings and the ways in which that understanding might be used to improve the teacher's work.

Panel 6: Teaching as Clinical Information Processing

This panel should be concerned with the thought processes that precede the teacher's choices of action. How teachers think, make judgments, size up situations, plan and react to unanticipated contingencies—all these should be of concern here. Thus, this panel would be concerned with the study of teaching as thinking, problem solving, judgment, and decision making. It might also consider the value of a clinical (possibly anthropological) approach to research on teaching. Studies of the teacher as clinician would be concerned with the way in which theory, although it does not prescribe for practice, nonetheless informs practice.

The diagnostic approach to teaching and teacher preparation would also be concerned with the teacher's expectations. Such expectations have been found to influence not the student's intelligence, as was once claimed, but teacher behavior, teacher-student interaction, student motivation, and student achievement of the objectives at which instruction is aimed. This panel should, therefore, give attention to the determiners and effects of teachers' expectations concerning students in various categories (sexes, races, ethnic groups, age levels, etc.).

Panel 7: Teacher Utilization

After teachers have been selected and trained, they go to work in a school which is part of a school system. What happens to them in the months and years of their work in the school obviously influences greatly the ways in

which they teach, feel and, what is most important, influence their students.

The term *utilization* refers to the ways in which teachers are required or permitted to work by the administrative, organizational, and physical environment that their school and school system provide. The many new ways of "utilizing" teachers in this sense include team-teaching, teacher aids, nongraded classrooms, schools without walls between classes, accountability laws, mandated teacher evaluation systems, voucher plans, teacher centers, flexible scheduling, and differentiated staffing. How is academic labor divided? What are the causes and effects of what is often considered to be a low level of collegiality in elementary and secondary schools? Should utilization variables be used in studies of school outcomes so that their value can be assessed? Can relationships between schools be arranged so as to improve teaching and inservice teacher education?

Sociologists, school administrators and teachers' organizations have been especially active in developing and analyzing innovations in teacher utilization. The panel here should contain representatives of these groups.

Panel 8: Personnel Roles in New Instructional Systems

Some writers regard the traditional classroom organization of the school and the teacher's role in that classroom as moribund. During the next two or three decades, the presently prevalent arrangements will be supplanted by what are now regarded as revolutionary new arrangements. The new schools will incorporate instructional systems that fully exploit the potentialities of computers, television, and various other products of recent technology.

The role of the teacher will change in correspondingly radical ways. In some views, the teacher will have much less to do with the intricacies of facilitating achievement of cognitive objectives, because computer-assisted instruction will carry much more of that burden. Instruction will be more effectively adapted to the needs and capabilities of each individual student.

Such individualization will adjust the pace, the occasions, the initial phases, the remedial efforts, and the media or methods of instruction so as to optimize learning of the individual student. Beyond these kinds of adjustment, the new systems will adapt the subject-matter presentations to the individual's aptitudes, learning styles, attitudes, preferred types of reinforcement, and immediate past history of responses. They will also use new approaches to the evaluation and grading of student achievement, turning more to "criterion-referenced" approaches. In turn, these approaches will lead to the objective of greatly increasing, indeed maximizing, the percentage of students that master all aspects of all their courses. Such "mastery" ideals will be achieved much more abundantly in the new instructional systems.

What will be the teacher's role in the new systems of instruction? Will the teacher become less the curriculum planner, the diagnostician and prescriber, the evaluator? Will the teacher become more group process facilitator, the mediator of emotional responses, the model for moral and esthetic development?

This panel will develop approaches to the research and development that will provide bases for predicting and optimizing the roles of various kinds of

persons in the new instructional systems. Basing its work on a variety of conceptions of the new systems, the panel will explore and formulate proposals for ways of making the emergence of the roles as efficient and rational as possible.

Panel 9. Research Methodology

As in most fields of the behavioral sciences, methodology in Research and Development on teaching, of the kinds with which panels 1-8 are concerned, has suffered from inadequate standards of validity.

The primary purpose of the panel on methodology is to identify methodological problems that appear crucial to the successful completion of research on teaching. The panel should produce a series of brief descriptions of problems that can be integrated into one component of a directed research program. The results of the directed research program component are to be methodological developments that are applied in the sense that they are of clear utility for research on teaching. It is also hoped that the directed research efforts will contribute to the more general field of psychometrics.

There are a number of methodological problems which need further attention and that relate directly to the value of research on teaching. In the area of measurement, there is a need for better understanding of the properties of existing measures of independent and dependent variables commonly used in past research. Such difficult problems as measuring the quality of a teacher (in several respects and across several situations) require new instrument construction. There is also a need for improved understanding of existing design and analysis strategies that would increase the utility of future research.

The Panel on Methodology will concern itself with both types of problems, with emphasis on the latter. In addition, it may call for a sampling survey of the methodological problems and errors in the recent literature on Research and Development on teaching. The finding of the survey might lead the way to preparing a set of methodological standards and guidelines for research and development in teaching. If they proved effective, such standards and guidelines would raise substantially the validity of the work of the next decade. They would also enhance materially the degree to which research findings would become comparable across studies, so that future reviews of the literature would be able to accumulate results in valid ways. Of course, at the same time it is necessary to remain sensitive to the ideal that methodology should be especially well suited to the particular investigation. Therefore, it would also be useful to develop methods for synthesizing results from studies having different methodologies.

Summary

This, then, is the strategy being tested by the NIE to increase the impact of the research on teaching. If the strategy is only marginally successful, we should move a long way toward providing the data needed to answer the question: "What are the things that teachers do which make a difference for children?"

3 Closing the Knowledge Gap*

H. Del Schalock

The position paper prepared by the Consortium of CBE Centers has three purposes. The first is to suggest the kinds of research needed to close the knowledge gap that now exists with respect to competency based teacher education. The second is to build a case for CBTE programs as unusually rich contexts for basic research in education. The third is to outline a framework for proceeding with basic research that will yield information of greater use to practitioners than have most basic research efforts in education in the past.

Background

The paper stems from the experience of the elementary model directors in attempting to design and implement CBTE programs in their own institutions, or in attempting to help others do so. In the course of these efforts, all have experienced the frustration of not having sufficient information on which to base program design decisions; all have recognized that there is little hope of improving the quality of what is being done until there is increased knowledge of what works and what does not work; and all have felt concern about the rapid move to adopt CBTE programs without clear evidence that what is being adopted is any better than what has been in opera-

* This is a summary of a position paper prepared by the research committee of the Consortium of CBE Centers on CBTE programs as a focus of and context for research in education. Members of the research committee are Dr. Vere DeVault, University of Wisconsin, Dr. Norman Dodd, Florida State University, Dr. Bruce Joyce, Columbia University, and the author. The paper is being published in full by the Multi-State Consortium on Performance Based Teacher Education. A comprehensive bibliography is included in that monograph.

tion in the past. The program of research that has been proposed by the Consortium of CBE Centers is designed to overcome these shortcomings, and in the course of so doing to move both education and teacher education closer to a mode of operation that is empirically based.

In addition to the pressing need to know, three other conditions have contributed to the decision to prepare a position paper on research in CBTE at this point in time. The first is that CBTE programs are reaching a level of maturity that permits the research called for to be carried out. Research on competency based teacher education, for example, or research that compares CBTE to non-CBTE programs, requires operational CBTE programs. Now that such programs exist, the essential condition for such research has been met.

The second condition that triggered the preparation of the position paper by the Consortium is the increasing demand for evidence as to cost and benefits associated with CBTE. Persons who control funds in support of teacher education programs and persons who have been led to expect significant changes in the educational system as a function of CBTE programs, are demanding to see data in this regard. Providing such data is seen as a fundamental obligation of the Consortium, as well as a necessary condition for the survival of the CBTE movement generally.

A third factor that has contributed to the preparation of the paper is the inconsistency that presently exists between CBTE model expectations and CBTE program implementation so far as a data based mode of operation is concerned. All of the elementary models called for continuous program adaptation and for this adaptation to be based upon empirically derived cost and effectiveness data. With few exceptions this basic tenet of the models has not been realized, and to maintain any degree of credibility—as well as to optimize the power of model based programs—it must.

Overview

The paper outlines four different kinds of studies that need to be undertaken if the knowledge gap with respect to CBTE is to be closed for all concerned. These have been labeled "Policy Oriented Studies," "Practice Oriented Studies," "Basic Research Studies," and "Documentation Studies."

As used in the paper, "Policy Oriented Studies" are essentially cost-benefit studies of CBTE versus non-CBTE programs, or of CBTE programs that differ in major respects, whereas "Practice Oriented Studies" are essentially cost-effectiveness studies of alternative approaches to the mastery of teaching competencies within the context of CBTE programs. The knowledge coming from practice oriented studies, of course, is as applicable to non-CBTE programs as it is to programs that reflect CBTE characteristics.

By contrast, "Basic Research Studies" are conceived as dealing fundamentally with two matters. The first is the search for new or more powerful constructs that can be applied to the field of education or teacher education. The second is the application of these constructs in the search for principles of instruction. As in the case of practice oriented studies, knowledge coming from basic research studies is equally applicable to CBTE and non-CBTE

programs

The proposal to carry out 'Documentation Studies' is in response to the need for usable case histories on the development and implementation of CBTE programs. Because of the relative complexity of CBTE programs and because of the general lack of familiarity with many aspects of their operation, help needs to be given to program designers and developers at a 'nuts and bolts' level. Documentary studies are seen as providing this kind of help. Three kinds of documentary studies are called for in the paper: one that describes the development of programs, one that describes the structure and operation of programs, and one that describes the diffusion of programs.

Because of space limitations, the research program being called for by the Consortium of CBE Centers is treated in the present paper essentially in outline form. This will be bothersome to some, for the level of detail needed to see what the research might look like when carried out is missing. To others, however, this lack of detail may make it easier to place in perspective the overall program of research that is being proposed. The reader who wishes greater detail than is offered in the present paper is referred to the Multi-State Consortium publication of the position paper in full (Schalock, 1975).

Part I: Policy Oriented Studies

Persons responsible for the allocation of resources in support of teacher education need to know whether the educational returns from CBTE programs are greater than those from traditional programs, whether the costs associated with CBTE programs are greater than those associated with traditional programs, and the relation of costs to benefits in both cases. Information of this kind is needed by persons responsible for establishing state and national standards for teacher preparation programs, by institutional executives who must allocate resources in support of programs, and by employees of the Federal government who must recommend or make policies relative to the funding of Federally sponsored activities that relate to such programs. Without such information, policy decisions must of necessity be made on some basis other than fact.

Two lines of research need to be undertaken in order to obtain information that facilitates policy decisions. The first involves the comparison of CBTE and non-CBTE programs as to costs and benefits. The second involves the same kind of comparisons, but of CBTE programs that define teaching competencies at different levels of generality, or at the same level of generality, but in markedly different ways. For ease of reference these two designs are referred to as

- a CBTE/non-CBTE design, and
- a CBTE Program Variations design.

Three conditions are essential to the implementation of either design. First, the criteria that define and differentiate desired program variations must be clearly specified. Second, ongoing teacher education programs must exist that meet these specifications. Third, persons responsible for

BEST COPY AVAILABLE

such programs must be willing to take part in a comparative research study. Put into other terms, the CBTE/non-CBTE research design calls for clarity as to what constitutes significant program differences, essentially "pure cases" of mature, fully functional programs that reflect these differences, and teacher educators willing to engage in a comparative research effort.

Given these conditions, policy oriented studies in the arena of CBTE will probably take the form of "natural" field experiments—that is, systematic comparisons of planned variations in already established programs.

CBTE/Non-CBTE Studies

The most critically needed policy oriented studies involve the comparison of CBTE and non-CBTE programs. Because there are major differences within CBTE and non-CBTE programs, however, and because these differences may be as significant in terms of cost/benefit relationships as the differences between CBTE and non-CBTE programs, an effort should be made to include within such studies a sampling of the major variations that appear within both CBTE and non-CBTE programs. On the CBTE side this should involve programs that define competencies primarily in terms of knowledge, skill, and the ability to perform the job of a teacher in a school setting.* On the non-CBTE side, programs should be selected that reflect the classic, discipline centered, four year undergraduate approach to teacher education, the "phenomenological" approach to teacher education, and perhaps the Master of Arts in Teaching (MAT), or the Teacher Corps approach to teacher education. These possible variations in program design are shown schematically in Figure 3.1

Figure 3.1

PROGRAM VARIATIONS THAT NEED TO BE TAKEN INTO ACCOUNT WHEN COMPARING CBTE AND NON-CBTE PROGRAMS

Dominant Variations in CBTE Programs	Dominant Variations in Non-CBTE Programs
1. Knowledge Centered Definition of Competency	1. Classic Four-Year Under- graduate Approach
2. Skill Centered Definition of Competency	2. Phenomenological Four-Year Undergraduate Approach
3. Job Performance Centered Definition of Competency	3. MAT or Teacher Corps Approach

In carrying out studies of this nature, it would seem that at least three programs would have to be found that meet the requirements of a particular program variation in order for that variation to be included in a study. Operationally, this probably means that the programs studied would center at

* According to Turner's criteria for classifying the focus of CBTE programs (Turner, 1972), knowledge-centered programs focus at level six, skill-centered programs focus at levels five, four, or three, and job performance centered programs focus at levels two or one.

the elementary level, for it is doubtful that a sufficient number of fully developed CBTE programs exist at the secondary level to permit the design requirements of three programs per cell to be met.

In addition to controlling as carefully as possible for program differences at the treatment level, the CBTE/non-CBTE research design requires that careful attention be given to the output or dependent variables to be considered, the unwanted sources of variation to be controlled, and descriptors of program cost. Table 3.1 contains a summary of preliminary thinking about such matters.

Program Variation Studies

This design is simply an elaboration of the CBTE portion of the design described previously. In the CBTE/non-CBTE design, it was proposed that CBTE programs included in a study vary by level of competency definition. In the present design it is proposed that *within level differences* be explored as well.

Operationally, the CBTE Program Variations design calls for the identification of programs that define competency primarily in terms of knowledge, skill or job performance, and for the comparison of major program variations within each of these three levels of competency definition as to the effectiveness of their graduates on the same criterion measures as employed in CBTE/non-CBTE studies. All of the conditions that need to be met in order to carry out CBTE/non-CBTE studies must be met in carrying out CBTE Program Variations studies, for example, mature CBTE programs of varying foci and emphases must be available, people in those programs must be willing to engage in a comparative study, and the same unwanted sources of variation must be controlled. The same dimensions of program cost must also be attended to.

Unfortunately, given the relatively limited development of CBTE programs nationally, it is likely that this design will be applicable in the near future only to programs focusing at the knowledge and skill levels. There simply are not enough CBTE programs in operation at the present time that define competencies at the job performance level to permit the comparison of the effects of alternative program emphases to be meaningful.

Part II: Practice Oriented Studies

While policy makers need the kind of information alluded to above, persons responsible for the development and implementation of CBTE programs need a different kind of information. Obviously, they must also be concerned with cost and benefit questions, but the central focus of program developers and managers is the effectiveness of alternative program structures and procedures in bringing about a specified set of teaching competencies. Examples of the kinds of information needed at this level of decision making include the costs and benefits of a field based, as opposed to a non-field based, mode of operation; the costs and benefits of "simulated" field conditions as opposed to "real-life" field conditions; the costs and benefits of implementing an instructional program in "module" form as

Table 3.1
OUTPUT MEASURES, SOURCES OF VARIATION TO BE CONTROLLED AND COST ANALYSIS
TO BE PERFORMED IN A CBTE/NON-CBTE COMPARATIVE DESIGN

Output Measures		Sources Of Variation To Be Controlled	Cost Analyses To Be Performed
Short Term	Long Term		
Job performance measures of the kind employed by Oregon College of Education in their assessment of competency in "short term" (2-5 days) or "extended" (2-5 weeks) full responsibility teaching	Job performance measures of the kind employed by Oregon College of Education in their assessment of competency in "long term" (2-5 months) or "continuing" (2-5 years) full responsibility teaching	Match between program design, program operation, and definitional criteria	Instructional costs
Other measures of impact on program operation, e.g., acceptability, organization and time requirements, resource requirements		Number of students graduated through the program	Competency assessment costs
		Level of preparation offered by the program, e.g., preparation for elementary teaching	Program assessment and redesign costs
		Kind of institution within which the program rests, e.g., private small college, private large university, public small college, public large university	Program management costs
		Geographic location	Program governance costs

opposed to a form that relies upon syllabi, readings, discussions, and the like; and the costs and benefits of including in a program elements that deal with understanding self as teacher, interpersonal awareness and sensitivity, and skills in group problem solving. Without such information, decisions as to program design and operation must also be based on something other than firm knowledge.

As in the case of policy oriented research, two conditions must exist before practice oriented research can be carried out. First, programs must be in operation that reflect the practice of concern. Second, either programs must be found that are comparable in structure, organization, and student body and are willing to compare alternative practices, or a program must be found that is willing to design and carry out alternative practices under experimental conditions while continuing to carry out program operations.

In some respects, these conditions resemble those needed to carry out policy oriented studies. The existence of fully operational programs is a case in point. In other respects, however, practice oriented studies go beyond policy oriented studies in their demand upon control over unwanted sources of variation. In this respect they become more difficult to carry out, for they impose considerably greater constraints upon programs that are taking part in the research.

Given these conditions, practice oriented studies in the arena of CBTE will probably take the form of "systematically manipulated" field experiments, in contrast to the "natural" field experiments that characterize policy oriented studies.

By viewing the requirements for practice oriented studies in this light, both policy makers and practitioners should be able to understand why research of this kind has not been undertaken thus far in the history of the CBTE movement. Simply put, it is for the same reason that policy studies have not been undertaken: there have not been programs sufficiently well developed to permit such research to be carried out. Now that CBTE programs are beginning to reach maturity, research of this kind becomes possible.

Intra-Institutional Studies

Since the effects of program practices and procedures would be expected to interact powerfully with learner characteristics and the kind of learning outcomes desired, it is likely that most practice oriented research will be carried out within the context of a particular CBTE program. This will permit as many sources of unwanted variation in such studies to be controlled as possible. The conditions outlined earlier, however, must be obtained before a CBTE program would qualify as a site for such research. Paraphrased, these conditions would include a fully operational CBTE program where personnel of the program are interested in and able to carry out experiments on the effects of systematically designed program variations while keeping constant the learning outcomes desired from the program. In addition, a participating institution must have a reasonably large number of students enrolled in its CBTE program; it must be organized in such a way that bona fide experimental and control groups can be established; it must be organized in such a way that true component variations are able to be im-

Table 3.2
OUTPUT MEASURES, SOURCES OF VARIATION TO BE CONTROLLED,
AND COST ANALYSES TO BE PERFORMED IN PRACTICE
ORIENTED STUDIES

Output Measures		Sources Of Variation to be Controlled	Cost Analyses to be Performed
Short Term	Long Term		
Whatever measures of teaching competency a particular program has adopted as the criterion for exit from either the program as a whole or segments within it (these may be developed as a by-product of participation in particular research studies)	None	Initial differences in treat- ment groups	Instructional costs
		Unwanted or unexplained treatment conditions	Competency assessment costs
		Lack of equivalent con- texts for competency demonstration	Program management costs
			Program governance costs
Other measures of impact on program operation, e.g., acceptability, organi- zation and time require- ments, resource require- ments			

plemented; and high quality measures of learning outcomes must be obtainable. The criterion or dependent measures to be used in such a design, the sources of variation to be controlled; and the dimensions of cost to be attended to are summarized in Table 3.

A note of caution needs to be introduced at this point. The fact that practice oriented studies require the careful control of unwanted sources of variation in both treatment conditions and outcome measures has led educational researchers historically to undertake practice oriented research in a particular way. First, it has tended to be carried out within contexts that would "tolerate" such manipulation—for example, an instructor's own class or some limited segment of a student population. Second, it has tended to involve definitions of educational practices that were relatively narrow or limited in scope, for example, a one day or one week program of instruction, a particular grading practice, or a particular set of role relationships between instructor and student that are to be tried for a term. Both strategies have given rise to the problem of the results of such studies having limited utility—even within the context where they were done—and limited generalizability.

These are the problems that have plagued educational researchers for decades. Only now are we coming to understand what must be done to overcome them. Basically they are problems that have to do with the external validity of education experiments. Shulman speaks to such problems pointedly when he says:

Researchers are caught in a bind. To maximize the *internal validity* of experiments, they develop carefully monitored settings within which they can govern their research. This has long been recognized as a necessity, but it is likely that the experimental tradition in American education over-emphasized the importance of reliability and precision at the expense of the characteristics affecting that other factor of equal importance (external validity) . . . It is not sufficient that the individuals studied as a sample are truly representative of that human population to which the results of a particular experiment will be inferred. Researchers must also ascertain that the experimental conditions can serve as a sample from which to make inferences to a population of external conditions of interest. That is, researchers must also attempt to *maximize the similarity* between the conditions in which they study behavior and those other conditions, whatever they may be, to which researchers may ultimately wish to make inferences. The similarity should hold between psychologically meaningful features of the settings, not merely between the manifest aspects of the two situations. (1970, p. 377)

If Shulman's analysis is correct, three fundamental shifts in the nature and focus of educational research must take place if it is to be responsive to the criteria of external validity: 1) a shift in emphasis from the sampling of people to the sampling of educational environments; 2) a shift from the

study of single, isolated variables to the study of the complex of variables that make up ongoing educational environments; and 3) a shift from piecemeal, unrelated, "one-shot" research efforts to studies that are articulated through time, theory, and problem focus. It is the contention of the research committee of the Consortium of CBE Centers that CBTE programs provide the kind of contexts where such shifts, for the first time, have some hope of coming about.

Having said this, it is important to point out that the conduct of long-term, multi-faceted, contextually valid experimental research lays heavy demands on an ongoing educational context, and unless the context is special in many ways, such demands will not be met. In the judgment of the committee, at least five conditions must prevail:

- *Persons responsible for the management and operation of the program must be inclined toward experimentation.* Commitment to empiricism and the desire to know must be dominant features of the context. Research must be viewed as an integral part of program operation, and as such viewed as a continuous, necessary, and desirable part of the program.
- *Persons responsible for the management and operation of the education program must view the program as subject to continuous change, and view as a major data source for its change a systematically designed program of research on its effectiveness.* When viewed in this way, research can have both the immediate and applied value that is needed for it to be supported by those responsible.
- *Data of a quality that will support trustworthy research must be collected as a normal part of program operation.* Accurate, reliable descriptions of program operations must be provided and accurate, reliable measures of learning outcomes must be recorded as a matter of course.
- *Sophisticated data management, storage, retrieval, and display capabilities must be available.* Data to be used for research, program operation, and program adaptation purposes must be routinely stored on computers and routinely retrieved in formats that support program related decision making and research.
- *There must be an advisory structure that ensures that the research that is pursued has value to persons in the program as well as to the profession at large.*

It is recognized that these five conditions represent an unusual array of features to find in any ongoing educational context today. It is also recognized that they are difficult conditions to bring about if they are not already in existence. It is the judgment of the committee, however, that CBTE programs represent the most hopeful development in the history of education for such conditions to be met.

Interinstitutional Studies

While most practice oriented studies will be carried out within a particular institution, at least initially, the continuing concern for the utility and gen-

eralizability of the results of such studies requires that they be replicated ultimately in a variety of educational contexts. To some extent, this requirement can be met by testing a particular practice in different programs, or with different populations of students within a single program. In the long run, however, this is not sufficient, for until it can be shown that a particular practice is successful in a number of institutions, it is not likely that it will have a great deal of acceptance in or utility to teacher education at large.

In carrying out interinstitutional studies on alternative practices, it will be necessary of course to control for the same sources of variation and to take the same measures of cost that are outlined in Table 3 2

Part III: Basic Studies

Historically, research in education and teacher education has been disappointing (Gage, 1963; Smith, 1972; Travers, 1973). A host of explanations have been offered for this fact, including the argument that the world of education is simply too complex to yield to the strategies of science, but two explanations seem to have most authority. The first is the failure of education research to meet the requirements of external validity. The second is its lack of adequate dependent or outcome measures. No matter how powerful a research design, or how powerful a treatment condition, research results of significance will not be forthcoming unless the measure of effect is strong.

One of the most hopeful features of the competency based education and teacher education movements is that they are designed specifically to overcome the perennial problem of outcome measurement. If carried to successful completion, competency based education should yield new and powerful measures of learning outcomes in pupils, and competency based teacher education should yield new and powerful measures of teacher performance. If these measures are of the kind and quality anticipated, and the research community recognizes them as such and takes advantage of their availability, basic research in education and teacher education should profit immensely.

Two conditions must be met, however, if competency based education and teacher education programs are to become contexts for basic research. First, as implied in the above paragraph, the measures employed in the assessment of teaching competence must be of a quality that permits their use in research. That is, they must be valid, reliable, and sensitive. Second, experimental designs must be employed in the context of program operation with sufficient rigor that "causality" can be attributed to the experimental or treatment variables investigated. Both of these conditions are above and beyond the requirements of normal program operation, but both can be achieved if introduced with care and foresight. When both are met, much is to be gained, for basic research can be carried out as an adjunct to normal program operation at little added cost.

While it is possible to combine basic research with program operation by meeting these two conditions, it needs to be pointed out that considerable risk is involved in attempting such a venture. High quality measures, for example, are often difficult and costly to obtain. Also, requiring that pro-

gram operations meet the constraints of experimental design almost always creates a clumsiness and rigidity that frustrates program managers and participants. Heretofore efforts to design data collection systems that support both program operation and basic research have tended to end in the design of research programs instead of operational programs that have good data. When this has occurred, there has been a nearly universal reaction on the part of program managers and participants throw the researchers out! (Parlett and Hamilton, 1972).

If this pitfall is recognized, it is still possible that, if done with care, data generation systems can be designed that will support both program operation and basic research. When this is the case, the best possible context for basic research exists: it can be carried out at low cost, and it has a good chance of meeting the requirements of external validity that are not met in most educational experiments.

Notes On Terminology

While the distinction between basic and applied research has never been particularly clear (Cohen, 1948; National Science Foundation, 1960; Brooks, 1967) the distinction has been maintained for purposes of the present paper. Following the proposals of Schalock and Sell (1972), basic research on instruction has as its focus (a) the identification and delineation of the *constructs* central to instruction and (b) the establishment of *principles of instruction* that incorporate these constructs. Applied research, on the other hand, has as its primary focus the testing of instructional practices, which may or may not reflect the principles of instruction developed through basic research. When viewed in this way, both the policy and practice oriented studies described in the previous sections of the paper are seen as classes of applied research. They have direct and immediate utility to practitioners and policy makers. Typically, basic research has little direct utility to the practitioner. Its primary utility is to the applied researcher, particularly the applied researcher who is concerned with practice oriented studies.

Another distinction that has never been particularly clear, but one that is central to the present paper, is the distinction between learning and teaching. As strange as it may seem, it is a distinction long ignored by educational and psychological researchers (Gage, 1963) but one that in recent years is beginning to bear fruit (Nuthall and Snook, 1973; McNeil and Popham, 1973). As used in the present paper, theories of learning deal with how an individual takes in, stores, transforms, and utilizes information; theories of teaching deal with how an individual influences that process. In drawing this distinction elsewhere, Schalock has pointed out that

In a sense the distinction is arbitrary, for the end point of both is learning. In another sense, however, it is not, for the focus of one is upon the *processes* of learning and the other upon the *conditions* of learning. By forcing the distinction, and then attending systematically to the conditions of instruction, issues that tend to be obscured when focusing upon learning come into full view; for example, the dependency of instructional decisions upon educational objectives and

learner characteristics. In their pursuit of laws governing learning, experimental psychologists have not attended systematically to either of these classes of information. . . . The assumption underlying the present effort is that by highlighting the instructional process, and by attending to it both conceptually and empirically, it will be possible in time to develop a productive science of instruction which, in turn, will permit an effective and efficient technology of instruction. (1972, p. 39)

A Paradigm For Basic Research On Teaching

Without wishing to become embroiled in the paradigm-model-theory debate (Brodbeck, 1963, Gage, 1963, Snow, 1973), it is necessary to present the conceptual framework that underlies the proposals to be made. Whether the framework is a model, a paradigm, or a first cut at a C- or D-Theory is unclear. What is clear are the directional implications of the framework for research on teaching. Since these directional implications are immodestly seen as " . . . defining the legitimate problems and methods of a research field for succeeding generations of practitioners" (Kuhn, 1962, p. 10), the framework is equally immodestly referred to as a paradigm.

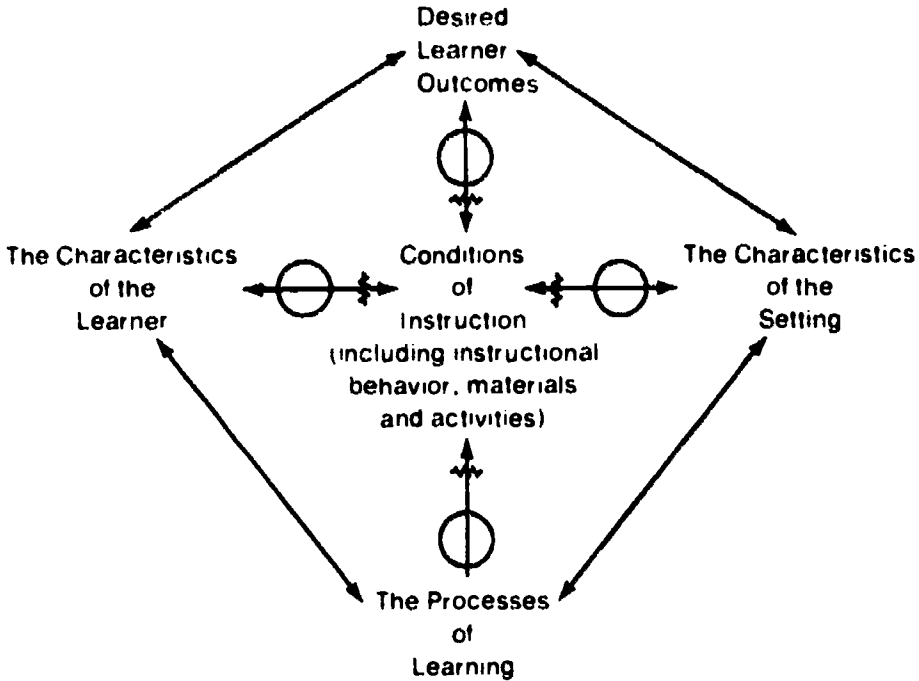
The paradigm proposed as a guide to basic research on teaching reflects a point of view about teaching that is not inconsistent with how teaching is viewed generally, but it is inconsistent with how most research on teaching has been done. The paradigm can best be described as a "context dependent" paradigm, for it is based on the assumption that for teaching to be effective, what a teacher does must always be adjusted to the nature of the learning outcome(s) desired, the nature of the learner(s) being taught, the nature of the setting within which learning-teaching is occurring, and whatever is known about how learning occurs. As such, effective teaching is seen as representing an articulation of a remarkable number of variables, especially when a wide variety of learning outcomes are being pursued by a wide variety of learner types under a wide variety of learning conditions. The paradigm is portrayed schematically as Figure 3.2.

Two comments need to be made about the paradigm. First, it has strong directional implications for research on teaching. It specifies the major classes of variables to be attended to in such research, and it specifies that these variables be treated in designs that accommodate them as interdependent or "contextually dependent" variables.

Second, it lends perspective to the anxiety that many educators feel about the continued emphasis in CBTE on *teacher behaviors* and the continued press to carry out research that attempts to link teacher behavior to pupil outcomes (see articles by Rosner and Kay, McDonald, and Flanders in the January, 1974, issue of the *Phi Delta Kappan* as examples of the continued press for such research). Within the framework of the proposed paradigm, teacher behavior may be treated as either a dependent or an independent variable, and when it is treated as an independent variable it must be done so within a design that links it to student characteristics, setting characteristics, class of learning outcome pursued, and the learning processes in-

Figure 3.2

FACTORS TO BE CONSIDERED IN ESTABLISHING
THE CONDITIONS OF INSTRUCTION AT A GIVEN
POINT IN TIME



(The circles on the arrowed lines represent learner behavior and the squiggles on those lines represent the filtering effect of teacher characteristics on the perception of learner behavior and the factors affecting learning.)

involved in the mastery of such outcomes.

Within such a framework, the continued search for simple correlations between teacher behaviors and learning outcomes in children does not carry much appeal. Nor does it carry much hope, considering the near infinite array of permutations possible between particular behaviors, particular learnings, particular learning outcomes, particular teachers, etc. The conclusions reached by Rosenshine and Furst (1971), therefore, should not be surprising. Nor should the kind of criticism that Maxwell has recently leveled at persons responsible for the design of CBTE programs who continue to define teacher competence in terms of specific teacher behaviors:

We don't have a list of the competencies of the successful teacher for the simple reason that teaching is an art—not a science. Behavior that is a valuable asset to a given teacher

with a certain personality in many (but not all) situations, insofar as most (but not all) of his pupils are concerned, may well go over like a lead balloon (to most students) if affected by some other teacher with a different personality. We can't even say that a teacher should love his students, since many teachers that we and the students consider successful exhibit no behavior (measurable or otherwise) that would lead one to this conclusion.

The teacher/student relationship is one characterized by heterogeneity on both sides of the relationship. The art lies in knowing, or sensing, what particular stimulus or behavior is best for me as an individual teacher to reach most effectively a particular student or group of students in a given set of circumstances. To assert that the "competent" teacher "will demonstrate his concern for persons in the lower socio-economic groups by voluntarily spending at least two hours per week working in a social agency" is absurd, no matter how highly we value such concern or approve of such work. We can make no such simple statement about the competent teacher (1974, p 310)

Research On Constructs That Relate To Instruction

In order to carry out research on instruction that meets the requirements of the paradigm outlined above, it is necessary to define, establish the parameters of, and make operational the major variables contained within it. The research and development involved in this process is perhaps the most basic of all activities that goes on in any science. It is a highly conceptual activity; it involves basic methodological work in that constructs must carry operational definitions and be subject to measurement in order to be incorporated into a research program; and it involves a major empirical effort in order to determine whether the constructs identified are sufficiently powerful to warrant inclusion in research on the paradigm proper. In its more advanced stages it is also likely to involve theory testing, since any set of constructs take their definition within a theoretical system of one kind or another.

For all these reasons, it is proposed that construct identification and measurement be considered as a necessary part of any basic research program on instruction and that it be seen as paralleling or preceding research on principles of instruction, that is, research on the relationships between the sets of variables outlined in Figure 3.2.

The design of construct identification and delineation studies assumes two basic forms. The first is a classic, single variable or comparative research design where the purpose is to determine whether a particular construct has any identifiable impact upon learning outcomes in children or has any likelihood of such impact even though there is no empirical evidence to that effect. Research of this kind is always guided by the paradigm that reflects the particular view of instruction that is being researched, for example, an "inquiry" or a "precision" approach to instruction. The specific pur-

pose of such research is to define within the broad classes of variables identified within Figure 3.2 specific variables that are promising for inclusion in the next level of research, that is, research that seeks to establish laws or principles.

A second research design that is a necessary feature of construct identification and delineation studies is a "testing of methodology" design. When new constructs are identified, or when existing constructs are refined, a set of measurement operations have to be developed in order to test the power of the constructs. Since the results of any research study are only as good as the measures of the constructs employed within it, the development of strong measures for all constructs to be investigated becomes a critically important activity. Central in this activity are comparative studies undertaken to determine the relative power or effectiveness of alternative approaches to the measurement of a particular construct.

Research of this kind, of course, borders upon and at times may become synonymous with the reliability and validity studies that must also be carried out as a part of methodological development.

As in the case of validity studies, the issue of criterion is paramount in the design of comparative studies of methodology. On what basis is the decision to be made that one approach to measurement is superior to another for a particular variable or construct? The correlation of one measure with another that is known to be a "valid" measure of the construct under consideration (the criterion used in "concurrent" validity studies) is an appropriate strategy, for it provides no information at all as to which of the two measures is in fact the more powerful. To obtain data on the relative power of a measure, one has to move to the arena of prediction. As tortuous and convoluted as predictive evidence often becomes, it is the only evidence that can be looked to with confidence in tests of the comparative power of alternative methodologies. Other things being equal, the measure that predicts with most accuracy or power must be viewed as the better measure (Schalock, 1969).

In closing the discussion of construct identification-delineation studies and the methodological studies that are a necessary part of such research, four observations seem appropriate. First, studies of this kind do not need to "start from scratch." Large literatures exist with respect to at least four of the five clusters of variables identified in Figure 3.2, and a literature is beginning to emerge in relation to the fifth. Second, research of this kind has not, and need not, be carried out only by educational researchers. Biologists, psychologists, anthropologists, and sociologists are all involved in the identification and delineation of constructs that contribute to the conceptual structure called for by the paradigm. Third, the process of construct identification and delineation is laborious, complex, and time consuming. It is also never ending, for implicit in the philosophy of science is the assumption that constructs will never cease to evolve nor ever find their absolute definition. In this context, it is probably safe to say that several hundred lifetimes of research will be needed on each of the classes of variables identified in Figure 3.2 before constructs will surface that are sufficiently powerful to permit research in education to proceed with the hopefulness of having an effect which research carries in the basic sciences.

Finally, the same reasons that make a CBTE context particularly desirable for carrying out practice oriented studies are the same reasons that make it a particularly attractive context for construct identification-delineation studies. It is a stable, continuously available research context; much of the data needed for such research are available as a consequence of normal program operation, thus making it a low-cost research context; and the variables being investigated are the real-life variables with which teacher educators and educators must deal. Stability, accessibility, low cost, and the likelihood of being able to overcome the problem of the external validity of experiments constitute a lot of value to be gained.

Research on Principles of Instruction

When construct identification-delineation research has progressed to the point where reasonable confidence can be placed in one or more variables within the various classes of variables specified in Figure 3.2, research on the identification of instructional principles can begin. As used in the present paper, a principle of instruction is an empirically established relationship between a particular set of instructional conditions and a particular class of learning outcome for learners who have a particular set of learning-related characteristics and are working under a particular set of learning conditions. The power of such principles rests in the guidance they provide for on-line instructional decisions made by teachers, and the more reflective decisions made by the designers of instructional materials and programs. If the empirical relationships established are strong, teachers and the designers of instructional materials will have a set of guidelines that for the first time will enable them to establish instructional conditions and be reasonably sure that those conditions will make the kind of difference desired. Until such empirical relationships are established, instruction and the design of instructional materials must of necessity be an iterative process of best guesses and successive approximations.

In some respects, the design of research that aims toward the identification of principles of instruction is more complex than that aimed at the identification of constructs and development of methodology, but in other respects it is less so. On the one hand, principle-related research requires the use of designs that will accommodate a large number of sources of variation in the course of a single study, whereas construct related research does not. On the other hand, the criterion problem is less troublesome in principle-related research than it is in construct related research. The criterion of progress in relation to desired learning outcomes is straightforward and unambiguous, whereas the criterion of predictive utility is anything but. Be this as it may, there are designs and methods of analysis that can accommodate the complexity required by principle related research. Prototypic in this regard are the designs and analyses used in aptitude-treatment interaction studies (Bracht, 1970; Cronbach and Snow, 1969).

As in the case of construct related studies, a number of brief comments need to be made before leaving the discussion of studies aimed at establishing instructional principles. First, studies of this kind do not need to "start from scratch," but they do not have the conceptual and empirical base enjoyed by construct related studies. Aptitude-treatment interaction

studies are prototypic, but they need to accommodate the additional variables identified in Figure 3.2. Second, research of this kind will probably be undertaken *only* by educational researchers. While biologists, psychologists, sociologists, and anthropologists search for and work with many of the same constructs that need to be involved in instructional research, they will rarely search for instructional principles of the kind considered here. Third, the process of establishing instructional principles is just as laborious, complex, and time consuming as the process of delineating constructs and developing methodology. It is also as neverending. In part, this is because of the interdependence of construct and principle related studies—as better constructs are established, better principle-related research can be undertaken; as better principles of instruction emerge, sharper constructs and finer measures can be developed. In part, however, it is because of the implicit assumption in the philosophy of science that principles once established will forever be refined.

Finally, the reasons that make a CBTE context a particularly attractive context for all of the other kinds of research that have been discussed are the same reasons that make it a particularly attractive context within which to pursue research leading to the establishment of instructional principles. It is a stable, continuously available context where the cost of carrying out research is low and the fidelity (external validity) of the variables high. Given the likelihood of a large number of "pilot" or "exploratory" studies being performed (the frequency of false leads and wrong predictions will undoubtedly be high in the beginning) and the requirement of a large number of replications to establish the confidence levels needed if practitioners are to rely upon the emerging principles as guides to instruction and instructional development, a research context that has CBTE-like properties would seem to be a necessary condition for the conduct of principle-related research. Whether ongoing CBTE programs can be found that will tolerate the demands placed upon them by such research, and at the same time possess staff with the competence to carry out such research, is an open question. Until such contexts are planned for, research that makes an appreciable difference in instructional design and practice is not likely to occur.

Part IV: Documentation Studies

It was argued earlier in this paper that CBTE needed to be the focus of documentary or case studies as well as basic and applied research studies. Competency based programs are a great deal more complex in their structure and organization than traditional programs and require for their development and implementation a level of assistance that new programs out of a traditional mold do not. The point of view adopted in the present paper is that an extremely useful way to provide such assistance, and indeed perhaps the only way, is through detailed documentation of the development and operation of existing CBTE programs. Knowledge of the development and implementation process in CBTE represents as much a knowledge gap as the knowledge gaps to be closed through the research directions that have been outlined. At least three kinds of case studies are needed within which both developmental and operational issues need to be attended to. One kind of case study needs to focus on program development and opera-

tion; another one selected aspects of CBTE programs that are particularly complex to develop and operate, or that have been particularly tardy in their development; and still another on the diffusion of program practices within an institution or to other institutions within a state or region. Documentations of the first kind would require finding programs reasonably well along in their development that reflect in their operation all of the basic principles called for in the generic conception of competency based teacher education. Documentations of the second kind would require only that the aspect of program operation to be documented be under development. Examples of program elements that might be the focus of second level documentations include selected consortium or coalition arrangements, selected approaches to competency definition and assessment, and selected approaches to program adaptation on the basis of program and student performance data.

Documentation of the third kind would require the identification of programs that could point clearly to program related practices or procedures that have been adopted by others. This type of documentation would allow for identification of diffusion related activities that could account for the adoptions.

Collectively, these various approaches to documentation would give the potential developer of CBTE programs access to critically needed information. They would provide access to strategies and procedures that have been used to develop such programs or to make such programs available to others. They would also provide some sense of the effectiveness of such strategies and procedures. In short, if carefully selected, they would provide much of the information needed to select from among options to program development, assist in carrying out program development, and suggest means for helping others adopt procedures and practices tested through program development. All are conditions that must exist if the principles of CBTE are ever to be adopted widely.

Within all three kinds of documentation studies, attention should be given to both developmental and operational issues. It is further proposed that, within both developmental and operational foci, attention be paid particularly to *the critical decision points encountered, the decision alternatives considered, the decision choice made, the consequences of the decision made and, to the extent possible, the likelihood of the consequences of alternative decisions considered*. It is information of this kind and at this level of detail, especially as it reflects the reality of real life operational constraints, that is most critically needed by those who wish to initiate or disseminate CBTE programs.

On the basis of extended experience in preparing case studies of ongoing research and development programs (Schalock, et. al., 1972), it is essential that the preparation of documentary studies be undertaken with as much planning, as much forethought, as much professional expertise, and as large a resource base as any of the applied research studies that have been called for. It is unclear, however, how best to actually carry out such studies. One procedure would be to contract with an agency to carry out all documentaries. This would involve the preparation of staff, the collection of data, the provision of case study reports, etc. The advantage of such a procedure,

as compared to a series of independently initiated studies, are efficiency, comparability of reporting formats, and reasonably good assurance of quality. The disadvantage is the lack of firsthand familiarity with the data to be reported and the dependence upon the recall of others to provide such data.

An alternative approach would be to provide only for the central coordination and management of the documentation effort at the Federal level, for example, establish specifications as to substance, format, etc., and train staff from the institutions whose programs are to be reported and have them serve as scribes or reporters of the development and implementation process as it is occurring. Either procedure would require a considerable funding base, a great deal of preparatory and design work, access to expertise in case study preparation, and a well designed management system to insure optimum quality at minimum costs. Without a clear recognition of the complexities involved in the preparation of the kind of case studies needed, there is danger that an effort to prepare such studies will be underfunded and understaffed.

A by-product of the documentation studies outlined, or perhaps a companion to them, is a series of case studies that document the development and operation of CBTE programs that serve as contexts for research. While relatively few CBTE programs will meet the criteria needed to function as research centers, it would seem to be important to include them in the overall documentation effort. Given the promise that such centers hold for educational research and given the nature of the relationships that must exist between instruction and research if a program is to serve as a functional research context, it would seem that documentaries of this kind would be of particular interest to an agency such as NIE. Since the creation of CBTE centers for research may be to education what the creation of laboratories was to the physical sciences, agencies such as NIE should understand all they possibly can about how such centers function.

Part V: Problems, Prospects, and a Plan of Procedure

Outlining the kind of research and development program needed in relation to CBTE is one thing. Implementing what has been outlined is something else. It seems appropriate, therefore, to close a paper of this kind with a brief look at the task of translating what has been proposed into practice and to offer at least tentative observations as to how that translation process might take place.

Some Anticipated Problems

It has been argued that only now is it possible to begin to undertake the kind of research and documentation studies that have been called for in the paper because only now are functional CBTE programs beginning to be established. It is simply not possible to compare the cost and benefits of CBTE with non-CBTE programs until CBTE programs are available for comparison. The same argument obviously pertains to using CBTE programs as contexts within which to carry out practice oriented research, basic research, and documentation studies.

Even though illustrative CBTE programs are now beginning to emerge, a host of problems remain before the kind of research and documentation called for in the present paper can be undertaken. Many such problems are referred to in the course of the paper; for example, finding CBTE programs that are willing to face comparison with other programs or that are willing to accommodate the demands of ongoing practice oriented and basic research studies. Other problems remain, however, and need at least to be identified. The more important of these would seem to be:

- Assuring the continued development and operation of competency based teacher education programs as demonstration and research contexts;
- Securing the funding level needed to carry out the research and documentation studies outlined in the present paper;
- Finding competency based teacher education programs that are at a level of development and possess sufficiently competent staff to carry out the kind of research and documentation programs outlined;
- Fixing responsibility for the long term coordination and management of the research and documentation program called for.

A number of these problems revolve around the availability of funds. This is obviously the case with respect to the research activities that have been proposed. Funding is also involved, however, in the continued development and operation of CBTE programs. Unless viable CBTE programs exist, most of what has been proposed is, in the parlance of the day, inoperative.

Some of these problems are not so much a matter of money as they are the availability of personnel who are competent to carry out the functions that will need to be carried out. In this regard, two broad sets of functions need to be performed—developing and maintaining a CBTE program that can in fact support the kind of research that has been called for; and actually carrying out the research and documentation activities.

Finally, a number of these problems are plainly and simply a matter of the National Institute of Education and the U. S. Office of Education agreeing to the importance of the kind of research and development effort that has been outlined; the probability of its short- and long-term contribution; the priorities of the proposed program as compared to the priorities of other programs that are being supported; the priorities of the various activities called for within the overall program; and the responsibilities to be assumed individually and collectively by the two agencies for the implementation, funding, and management of the program. In approaching the various problems and issues that have been identified, it is recommended that the last issue mentioned be attended to first, since decisions reached at that level will largely determine what happens in relation to all else.

Some Optimistic Projections

Assuming that the National Institute of Education and the U. S. Office of Education were able to reach the decision that the proposed program was worth supporting, that it had high priority in relation to other programs, that agreements were able to be reached in terms of shared responsibility

for supporting and managing the program, that a reasonably high level of funding was made available to the program, and that operational CBTE programs could be found that were willing to serve as research contexts, what could be expected to emerge from the program at the end of a 3-year period? at the end of 5 years? at the end of 10 years? The projections that follow attempt an answer to these questions. They assume that all lines of inquiry that have been recommended are to be supported.

***Products And Conditions At
The End Of Three Years***

DATA FROM POLICY ORIENTED STUDIES

- Preliminary data on costs and benefits of CBTE and non-CBTE programs
- Preliminary data on costs and benefits of alternative forms of CBTE programs
- Methodologies, settings, and personnel established to carry out long-term policy studies

DATA FROM PRACTICE ORIENTED STUDIES

- Preliminary data on the costs and benefits associated with selected instructional practices within particular CBTE contexts
- Methodologies, settings, and personnel established to carry out long-term studies of the costs and benefits associated with selected instructional practices within particular CBTE institutions
- Methodologies, settings, and personnel established to carry out cross-institution replications of cost and benefit studies on selected instructional practices

DATA FROM BASIC RESEARCH STUDIES

- First level taxonomic work completed in relation to the construct delineation studies, related literature reviews completed, design and measurement methodologies tested, and some replication data on constructs that have appeared promising in existing education and psychological literature
- Theoretical positions that link the constructs identified in the research paradigm stated at an initial level of detail, related literature reviews completed, design and analysis methodologies tested, and studies in progress that test the soundness of the theoretical structures that guide the assembling of variables

CASE STUDIES FROM THE DOCUMENTATION EFFORTS

- At least three case studies on selected CBTE institutions covering at least two years of development and implementation decisions
- At least six case studies on selected attributes of competency based teacher education that have 2-year history of development and implementation

DATA FROM POLICY ORIENTED STUDIES

- Sufficient data on costs and benefits of CBTE and non-CBTE programs that policy decisions can be made with reasonable confidence
- Sufficient data on costs and benefits of alternative forms of CBTE programs that policy decisions can be made with reasonable confidence
- Methodologies, settings, and personnel established to carry out whatever additional policy studies are needed

DATA FROM PRACTICE ORIENTED STUDIES

- Sufficient data on the short-term costs and benefits associated with selected instructional practices in particular CBTE contexts that program implementation decisions can be made with reasonable confidence
- Preliminary data on the long-term costs and benefits associated with selected instructional practices in particular CBTE contexts
- Initial replications across institutional settings of cost/benefit studies on selected instructional practices

DATA FROM BASIC RESEARCH STUDIES

- An initial set of empirically grounded constructs within each cluster of variables identified in the research paradigm that has been proposed
- Measurement methodologies firmly established for each construct within each cluster of variables in the paradigm
- The theoretical structures that link the constructs identified in the research paradigm reasonably well worked out and reasonably productive in terms of hypothesis generation
- Design and analysis methodologies firmly established for research on principles of teaching, given the constraints of the paradigm governing principle-related research
- A beginning collection of empirically based principles of instruction that are suggestive of the nature of the interactions between the variables listed in the paradigm on teaching

CASE STUDIES FROM THE DOCUMENTATION EFFORTS

- Completed case studies on three selected CBTE institutions (each case study would cover a 4-year period of time)
- Six additional case studies of selected attributes associated with competency based teacher education programs that have had at least a 2 year history of development and implementation (making a total of 12 such documentaries)
- At least three case studies covering a 2-year period of time on institutions that are serving as centers for research in CBTE

DATA FROM POLICY ORIENTED STUDIES

- The completion of whatever cost/benefit studies are needed on CBTE as opposed to non-CBTE programs
- The completion of whatever cost/benefit studies are needed on alternative forms of CBTE programs

DATA FROM PRACTICE ORIENTED STUDIES

- Short-term cost and benefits studies essentially completed on selected instructional practices within particular CBTE contexts
- Sufficient data on long-term cost and benefits of selected instructional practices in particular CBTE contexts that program decisions can be made with reasonable confidence
- Sufficient cross-institutional replication data on cost/benefit studies of selected instructional practices that program decisions by a wide variety of institutions can be made with considerable confidence

DATA FROM BASIC RESEARCH STUDIES

- Second generation construct delineation research underway and the theoretical structures within which constructs rest under empirical test
- Second generation design and methodology in relation to construct delineation studies under development
- A sizeable number of principles of teaching established, and established with sufficient confidence that program design as well as on-line teaching is able to benefit from their application

CASE STUDIES FROM THE DOCUMENTATION EFFORTS

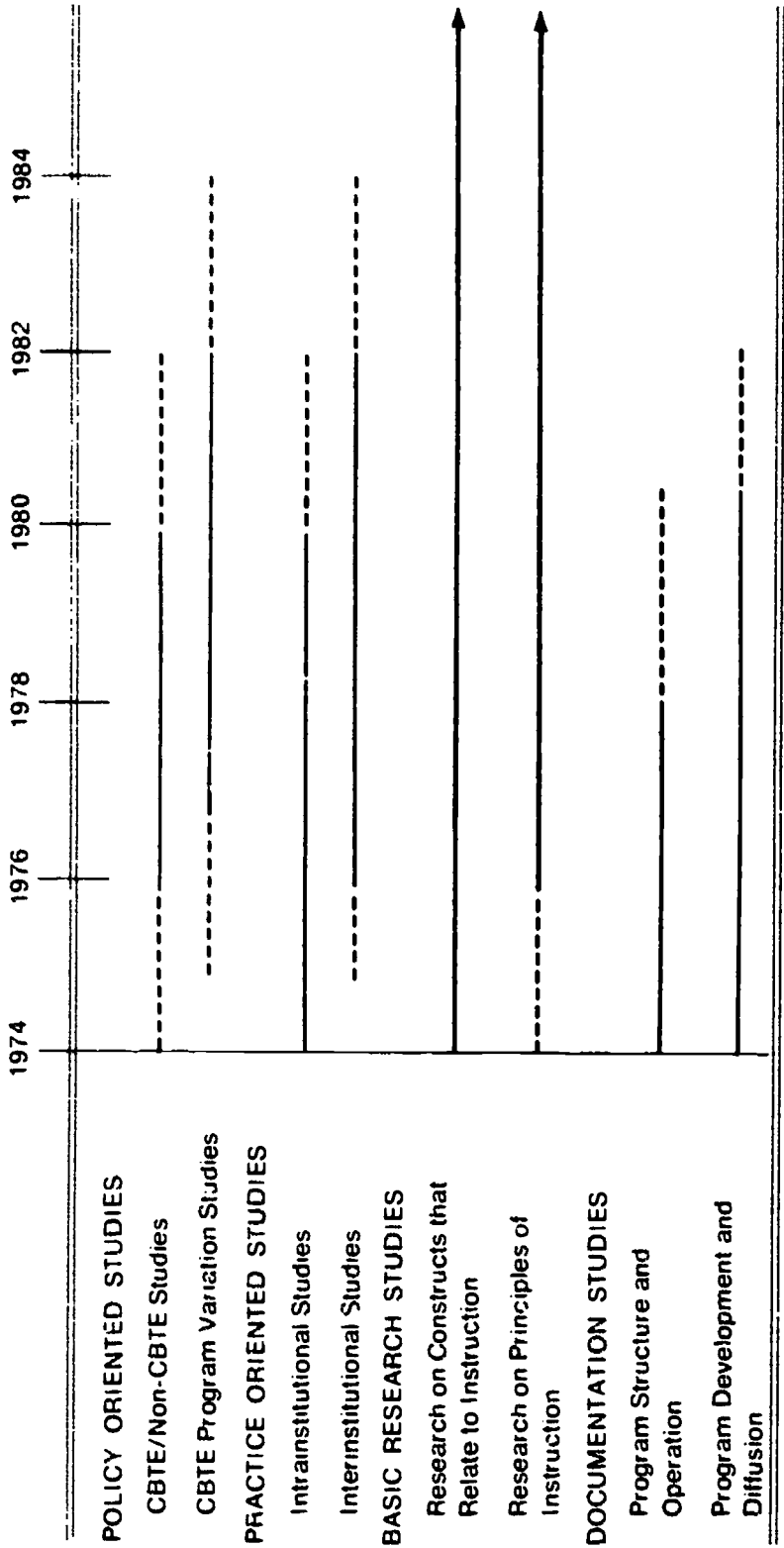
- All case studies on CBTE programs and program elements completed
- All case studies on CBTE programs as contexts for research completed

Figure 3.3 represents a translation of these projections into a simple time/task chart.

A Plan Of Procedure

Any of a dozen strategies could be followed in implementing the program of research and documentation that has been proposed. Each would have advantages and disadvantages; and more than one would undoubtedly prove to be successful. Be that as it may, a number of recommendations are offered at this point to assist in its implementation. The recommendations assume NIE and OE agreement as to joint sponsorship of the proposed program and agreement as to the division of responsibilities therein.

Figure 3.3
A TIME/TASK CHART THAT IS BEING PROPOSED
 (Solid lines represent periods of major research activity, broken lines represent beginning or closing activities)



BEST COPY AVAILABLE

1. Begin work on all aspects of the research proposed, even though there may not be enough money to do so on a large scale. All of the research efforts outlined need to be done; each will contribute to the other; and each will serve the needs of different audiences.
2. Take fully into account the program development work needed in order to carry out the research that has been proposed, and then see to it that the development and the research aspects of the overall program are closely coordinated. Without such attention to development there is danger that the research contexts on which the entire proposal is predicated will not be available.
3. Share the research and development costs across Federal agencies; for example, have NIE assume the costs of research and OE assume the costs of development and documentation.
4. Keep responsibility for managing the overall research program a Federal responsibility, rather than a state or non-Federal responsibility.
5. Hinge the program on competitive proposals that are in response to RFP's.
6. Direct most of the funding for research to institutions that have CBTE programs that can function as Research and Development contexts (funds should not be directed to existing Research and Development Centers unless they are linked to CBTE programs that meet the criteria for CBTE research centers).
7. Before letting research contracts to any institution, establish carefully the characteristics needed by an institution to carry out the various research and development activities that have been called for.
8. Implement a dissemination function of major proportions, and target the research and documentation results to be disseminated to the audiences for which they are particularly appropriate.

Obviously, the research program proposed is a massive undertaking. It is also an extremely complex one, and one that in the minds of many would be too slow to produce results or too grandiose ever to get started. In the judgment of the members of the Consortium of CBE Centers, however—and it is a judgment that has been carefully weighed—it is precisely this kind of research program that is needed in order for teacher education programs ever to make the kind of difference in teaching that they are supposed to make. The irony and hope is that by using CBTE programs as contexts for research, costs will be sufficiently reduced and the problem of the external validity of experiments will be sufficiently overcome so that the program that has been proposed can be carried out successfully.

BEST COPY AVAILABLE

PART II

ASSESSING PROFESSIONAL COMPETENCIES

4 Listening to Different Drummers: Evaluating Alternative Instructional Models

Bruce R. Joyce

The purpose of research on teaching effects is to identify teaching skills which induce learning of different kinds. Therefore, it is prescriptive research in the sense that it seeks relationships which can guide teachers, teacher trainers, and policy makers. By now most of us recognize that valid prescriptions will be developed only slowly and after intensive cumulative research programs have been established for some time. These programs will gradually clarify the outlines of skill in teaching and relate them to growth by learners. What we will find depends largely on what we learn to look for.

In this paper, I seek to clarify several issues which affect the shape of such research programs—their direction and their technology.

Some of these issues are technical and some are philosophical. We begin with a philosophical issue with considerable implications for the shape of our research.

The Teacher and the Learner: With Whom do we Start?

I firmly believe that we have to begin with the conception of the learner rather than a conception of teaching. I think we are in some danger of forgetting one of the most important messages that we can get from John Dewey's attempt to define the process of education.¹ Specifically, I am referring to the belief that learning is a product of the interaction of the learner with an environment. The importance of this definition lies in the emphasis on the learner's activity and the perspective that the educator facilitates learning by creating environments.

When we begin with the conception that it is the student's own activity

that results in learning and that the role of the educator lies in the provision of the environments with which the learner interacts. Then the teacher becomes important as he contributes to, and is part of, an environment. He is obviously only one of the contributing agents, but he is a critical one. His role broadens from an instructional one to the multidimensional set of roles that contributes to the creation of the environment (The teacher helps to build the social system, to develop a sense of rapport among the students, to bring resources together, to help the learner focus his activity, to set out tasks, and so on.) Teaching skill from this perspective is the creation of environments. This conception of teaching has a powerful effect on the search for skill and an even greater effect on the evaluation of the effects of a skill, for a skill appears as part of a complicated environment with which a complex bundle of sensations and capacities—the learner—interacts.

To begin with, we do not expect the effects of an environment to be mechanical and simple. A change in environment will ordinarily be expected to have only a gradual effect on the learner. He has to adjust to the environment, and if it is very different from previous ones, the adjustment usually takes some time. In addition, because it is the learner's activity that results in learning, his motivation, intelligence, experience, and skill will naturally be very powerful determinants of what results. We do not expect that environments will have a mechanical, homogeneous effect on any group of learners that interacts into them. At times, as I read some of the literature on competency based education and review some of the studies that are presently going on, I feel that the teacher is being conceived almost as a swordsman who lays about the learner with his repertory of sharp-pointed skills, each of which will have a mechanical effect on learning outcomes. The expectation seems to be that if he pushes on the learner with a certain kind of question, then a certain kind of learning will pop out, and we will be able to see it quickly and clearly. Such is not the nature of the interaction between learning environments and learners. It is very rare that any one skill (any single sword thrust) or any stylistic characteristic will stand out as the cause of the learning. Even where we find that increased learning is associated with the presence of a certain kind of skill or style in teaching, it is very likely that the reason is because that skill or aspect of style is an index of a much larger complex of behaviors that signals the presence of a certain kind of environment (For example, environments characterized by praise are believed by many to produce greater learning, certainly greater self-esteem. The skill of praising helps to create such an environment, but it is not very likely that it stands alone. A teacher who authentically and warmly praises his learners more than another teacher probably also does many, many other things as well. You do not simply turn on praise in a matrix of hostility and bring it forth as an unintegrated skill that shines out and determines what the learner will do.)

Teaching, thus, is a multidimensional act that fits in a multidimensional context to produce multidimensional environments which are interacted with by multidimensional learners. Given this condition, research on teaching effectiveness will produce a multivariate kind of truth. We will not end up with formulas that we can state in unequivocal, mechanical terms such as "teaching skill such-and-such increases such-and-such a kind of learning

in such-and-such a way. A teaching skill will be more likely a small but useful part of a certain kind of environment which will be reacted to differentially by learners and will have a number of kinds of main and side effects.

It is because of the really great complexity of the learner-teacher relationship that the study of teaching effectiveness appears to be a stab at a seamless web. To develop a cumulative program of research to identify and validate teaching competencies, a number of issues have to be faced. Some of these issues are philosophical, and some are technical. In this paper, we will discuss several of them and suggest how their resolution leads to decisions on the shape of a program of research on teaching effectiveness.

Cutting into the Web: How should we partition the Act of Teaching?

Teaching can be viewed as a generalized role-function—an organic activity whose variations we try to clarify and try to relate to learner activity—or it can be partitioned in terms of a variety of sub-roles and dimensions of activity related to those roles.

Gage² has severely criticized treating teaching as an undifferentiated function and attributed the slowness of growth in reliable indicators of teaching effectiveness to the lack of partitions in the definitions of teaching which have guided so many investigations. This has reduced the likelihood of identifying relationships between teaching skills and pupil behavior. For example, questioning behavior by teachers is often studied and related to pupil outcomes. We have to question whether any given aspect of teaching behavior is likely to be effective through all the roles the teacher plays, or without considering the complex of behaviors in which the targeted one is embedded. It appears to me to be very unlikely that any single teaching skill or even a combination of a number of skills is likely to be related very consistently to any kind of learning outcome except within a specifically defined role. Certain teaching skills may function beautifully in certain kinds of teaching roles and be totally irrelevant to others. Thus, to use the undifferentiated act of teaching as a dependent variable of training studies or as an independent variable with respect to all classes of learning outcomes in correlational studies seems to me to be unwarranted. The thinking which underlies most such studies is simplistic. The investigator tends to get into a frame of reference where he hopes to identify easily definable skills that are likely to be related to specific kinds of learning outcomes and to search for a simple cause and effect relationship among these.

I prefer to see teaching as an act which contributes to the creation of various learning environments which fulfill certain kinds of roles in the lives of students, roles which have attached to them different kinds of potential benefit for the learner. In other words, it appears to me that teaching is the creation of a variety of learning environments from which learners can benefit in a variety of ways. In another paper for example, I have described a school in which learners visit several kinds of centers, each one of which is characterized by an environment that can help him grow in particular kinds of ways. In the idiosyncratic center he meets counselors who help him to

find learning activities which will help him to grow on his own terms. In the skills center he meets with diagnosticians who help him understand his own competence in reading, mathematics, and some other "basic" areas and define with him programs of study to remedy his deficits and to help him obtain further skills. In the academic inquiry center he works with people who help him try ideas from the disciplines to learn the modes of inquiry that characterize several of them. In the social ecology center he works with others in social action developing the skills of active citizenship and learning to inquire into social problems and clarify the value issues of his society. Each center tries to foster a particular kind of independence in the learner—to teach him how to learn in a certain fashion.

Learning Center by Function in a Hypothetical School³

Center	Idiosyncratic	Basic Skills	Academic Inquiry	Social Ecology
Function	Personal Growth Program	Mastery of Symbolic Skills	Concepts and Methods of Disciplines	Valuing - Social Action
Role of Teaching	Counseling	Diagnostic - Prescriptive	Inductive	Group Inquiry

For our purposes here, the importance of this illustration is that each of these centers is characterized by a particular kind of learning outcome for which certain kinds of learning environments or modes are likely to be more appropriate than others. The teacher, the most important agent in creating those learning environments, probably needs different skills to work in each center. How much teaching skills would overlap from center to center is an open question.

In today's schools we tend to employ teachers as generalists, asking them to play a variety of roles analogous to those in the hypothetical school. A self-contained classroom teacher, for example, is asked to teach skills and concepts from academic disciplines as well as to counsel his students and to lead them in social dialogue. Just as in the differentiated staffing plan of the hypothetical school, each of these roles (and we might mention others) requires a range of unique competencies embodied in one teacher. Ultimately, a program of research to define and validate competencies needs to take into account these role differences. Particular kinds of questioning skills, for example, are appropriate to each model of teaching or to the environment which is most likely to produce the outcomes associated with each role. Many skills which may have important utility in one role may scarcely be relevant to another.

I believe that our collective effort to conceive of and validate teacher competencies depends on our ability to develop common definitions of teaching which partition the act so that we can search within defined fields. I believe that this was the major importance of Gage's recent article on the

microcriterion approach, where he suggests that the general acts of teaching are far too broad and ill-defined to permit systematic programs of research; but that if we identify particular kinds of functions of teaching such as lecturing and counseling, we create a more rational field within which to develop programs of inquiry.

For my own work, pending a dialogue among the group of researchers working in this area, I am searching within the roles of the teacher which are defined in Appendix B. I am attempting to identify the models of teaching which logically seem most productive for those roles: testing the models, and trying to locate the competencies that appear necessary to bring those models into reality inside of those roles.

The Philosophical Side: The Avoidance of Technophilia

The definition of teaching and the selection of environments are more than technical processes—we cannot study teaching as if it was an act without moral significance. For example, any model of teaching (the specifications of an educational environment) represents a philosophical statement. That is, it emerges from a view of man—how he learns, and what he should learn. Consequently, it is a symbol as well as an instrument in the life of the learner. The meaning of that symbol is likely to be as important an outcome of instruction as is any substantive or attitudinal learning which is likely to derive from instruction.

For example, the use of drill and practice strategies in military training symbolizes the discipline of the military and an obedience orientation. It represents a view of man as one who should be disciplined to collective purpose, who should learn to play his roles with precision and should confine himself to roles defined by an authority which is above him and in control of most of his behavior. The environment of Summerhill, by contrast, symbolizes the importance of the individual's psychosexual development and the necessity for him to define life on his own terms and to place his needs for personal integrity and fulfillment above those of the normative requirements of his society. He has to negotiate the demands of his own needs with those of his fellows, rather than have the means of that reconciliation thrust upon him.

In the search for the models which are appropriate for the various roles of the teacher, and indeed when we select those roles in the first place, we need to consider the symbolic effects and implications of our choices. If we search only for the most efficient methods, in the sense that we select only those models that result in defining learning outcomes in relatively short periods of time, we are making a philosophical statement as well as choosing an effective avenue. To avoid the implications of the symbolic value of the environments we select would be to beg the question of the nature of our existence and its meaning. At this time when there is so great a concern with accountability and when accountability is translated into a search for efficiency-related competencies, we must not forget that we create philosophical meaning in all of our actions. Because of this, to fail to research the symbolic effects of a model of teaching is not justified. I am suggesting explicitly that one of the necessary dimensions of our re-

search is to begin to understand the learner's interpretation of the symbolism inherent in our educational environments and the kind of philosophical views which he is likely to derive by exposure to them. We will presently return to the technical side of this aspect of research. As an issue, I place it early because the meaning of the remainder of our work is contingent on our handling of the symbolic worth of the types of teaching we choose to study and advocate.

Potential Pluralism: Does our Research Converge Toward a Monolithic School or Does it Diverge Toward Educational Pluralism?

I will not pretend to be open minded on this issue. The question is plain—Do we hunt for a single, all-purpose model of teaching, or do we try to find the range of competencies that are appropriate for an educationally diverse world?

The encouragement of diversity among mankind requires diversity in the nurturing of the young. A pluralism of educational environments is important for two reasons. One of these is because it exposes to the learner the operationalization of a variety of points of view about the nature of man and his development. It symbolizes pluralism and shows him what it is. Second, exposure to different environments is likely to promote different kinds of learning outcomes and thus increase the learner's repertoire of knowledge and skills. Our research should anticipate a much greater degree of pluralism than we see today in most of our schools, or certainly have seen in our past. Pluralism, it has to be noted, is different from eclecticism. I do not believe that we should search for a model of teaching which is a compendium of the most attractive elements of a variety of models. We should seek rather, I believe, to develop models on their own terms, borrowing only where it is clearly appropriate. And rather than searching for the optimal teaching strategy for all purposes, we should search for ways of creating the pluralism of environments that will lead to a pluralism of human kind. Our research should be relevant to today's schools, but not confined by them.

Behaviorism and Conservatism

A few years ago Michael Apple⁴ wrote a concise and pungent essay in which he pointed out that a search for teaching competence which relied largely on the study of teachers working in today's schools would be bound to have a restraining effect on educational progress, to put it mildly.

I have a great love for teachers and still see myself as one. Teachers labor, however, under ungodly conditions. Elementary teachers are best described, from a behavioral point of view, as able to teach any learner anything under any conditions (to achieve any objective with any learner using any teaching strategy). In response to their role-demands they develop an eclectic, elastic style that is anything but a repertoire of ultra-sharp competencies. If we have learned anything from studying them, it is that the demands of their role have funneled them into a style much more

that of a bureaucrat of learning materials than a master of technique.

The time-confines of the subject specialists, straightjacketed by the 45-minute class period, general course syllabuses, and adopted textbooks, have homogenized them also into the skills demanded by the four moves described in Arno Bellack's *Language of the Classroom*,⁵ the skills of the recitation. Most *Models of Teaching* do not find the customary role of the teacher or the skills of this role congenial to their execution.

If we finally succeed in identifying the competencies most useful for today's teachers, they will look more like the skills of a country postmaster than those of the creative artist or behavioral technician. Our task is, rather, to find the skills that will increase the capacity of tomorrow's teachers to vary—to discover the competence necessary to produce learning environments rarely spawned in the present, so-called "Real World" of the classroom.

If we can trust our theoreticians and researchers at all, we presently have a store of models of teaching which require competence rarely seen in today's classroom. (For one example: a student teacher and his students produce, in one hour of microteaching practice with the Synectics Model of Teaching, more synthesis-level communications than 30 experienced teachers produce altogether in 300 teaching episodes.⁶)

The Matching Models Concept: Building the Equation from the Student.

In addition, learners respond differently to any model of teaching. We should not expect any teaching skill, let alone any model, to have the same effect on all students. Without a differential concept there will probably never be such a thing as a validated teaching skill. What we may have ultimately is an elaborated matching models concept (such as the one proposed by David Hunt), in which the validation of a teaching strategy is reported in terms of its differential effect on a variety of learners. In other words, a program of research to validate teaching competence needs to anticipate the differential effects of environments on learners and build that capability into their research designs. Our empirical experience in the subject areas should reinforce this concept. The vast amount of research in reading methodology, for example, has shown very small differences among competing methodologies, except when one considers the characteristics of the learner. In other words, competence in the teaching of reading has to be conceived of partly as the competence to select a model that is most appropriate for the growth of the particular learner.

We may even find that the effectiveness of any given strategy may "wear off" after a period of time. The learner as well as the teacher may need variety to keep stimulated, and as he matures, his vacillating preference for environments may be very important to him.

The Settings of Teaching: The Context of the Roles.

When one conceives of teaching as the exposure of a learner to an educational environment, he immediately perceives that the agent is not the

only important element in the creation of the environment. Setting becomes extremely important to both the learner and the teachers. In recent years in our work at the Agnes Russell School, we have created a variety of learning centers which support different kinds of teacher activity. In the skills center, for example, the teacher works with the assistance of a variety of diagnostic devices and a range of systems which can be used by learners to remedy particular kinds of skill difficulties. This teacher in the skills center is facilitated in the role of skill builder by these support materials. In addition, he is affected by the concentration of his efforts, which has enabled him to train himself in specific role-related skills. He can use programmed instructional methods and other systematic teaching methods with a far greater accuracy and fluency than can some of the other teachers in some of the other settings. Similarly, the teacher in the Sciencing Center is supplied with a storehouse of equipment which students can use to replicate science experiments. The presence of that equipment, as well as the teacher's competence in using it, affects strongly the kinds of environments that can be created within that Center. Students can generate experiments, organize their equipment, and have places where they can keep the material for long periods of time while they're conducting a series of investigations

The Film Center was supported with videotape equipment, motion picture cameras, still cameras, and a darkroom. The teacher thus supplied was able to create an environment in which children learned the skills of filming and also could engage in production of films and television recordings.

Not only were the physical settings important, but also the setting of the Centers in the community of the school. For example, the Film Center worked on a contract basis with small groups of children. The Skills Center, on the other hand, was responsible for the diagnosis of reading and arithmetic difficulties for every student in the school. Appointments at that Center were mandatory for all children. The Sciencing Center negotiated with teams of teachers and their children to generate units of activity, as well as contracting with small groups of children. The possibilities of teaching are greatly affected by the fact that these Centers relate to their clients by appointment and by negotiation, as are the competencies needed to operate in the Centers. Competence is enormously related to both the material and social aspects of setting. The settings in many schools are so primitive that probably only a few kinds of educational environments can be created in them. The secondary teacher, working in 45-minute periods in a departmentalized organization has brief, if regular, contact with children. The roles he is able to generate in that kind of setting contrast greatly with those of the teachers in the Centers alluded to above. We need to increase knowledge about how setting influences competence, including how it is possible to activate different kinds of learning environments in a variety of kinds of settings.

Teaching as a Boost to Learning: Self-Directed Learner Activity as the Baseline

When one conceives of teaching as the creation of learning environments, the emphasis moves toward the learner. If learning is seen as a product of

the learner's interaction with the environment, then it is primarily the learner's activity which produces the learning. A teacher as an agent can help shape the environment and modulate it to the requirements of the individual child, but in the final analysis, there is truth in the old cliché that learning is done by the learner. Teaching can but facilitate learning, even in the most directive of the Models of Teaching. Because of this, the effects of a teaching competence—the extent to which it contributes to learner growth—needs always to be measured against what the learner might do for himself. The results of this can be somewhat surprising. David Hunt and I and our associates have conducted a series of studies in Toronto in which Marsha Weil and I have created environments styled after several models of teaching.⁷ We have exposed to those environments students of varying personality structure. We have compared the substantive learning (recall of information and the integration of subject matter) of students exposed to those environments with that of children who read the material for themselves and conducted their own analysis of it. Even after five or six hours of teaching we failed to raise recall scores above those of the "do-it-themselves" control group. After as little as one hour, however, we affected the numbers and types of concepts the children formed—the integration of material. In other words, we as teachers and the environments created by the models of teaching have a much greater impact on the learner with respect to "higher order" outcomes.

Studies of the effects of teaching should be measured against the effects of self-directed learner activity. (The same principle applies in teacher education; training strategies should be measured against teachers' self-training.)

The Multidimensionality of Educational Environments

Educational environments consist of a variety of cognitive and social tasks set in a social context and a support system of materials. The learner thus is exposed to a social system, some kind of sequence of social tasks, sets of cognitive tasks, learning materials, and the personalities of the individuals in the environment. The study of the effects of instruction is the study of the results of being exposed to a many-sided environment. We can expect multidimensional environments to have several kinds of effects resulting from a combination of the substances into which one inquires, the cognitive and social tasks, and the social relations.

An instructional model or model of teaching is a pattern for the design of such an environment. In practice, inquiry into the effects of models of teaching is complicated by the fact that so many instructional strategies do not explicitly attend to a number of the important dimensions of the environment that they will inevitably create. We have to flush them out by careful conceptualizing and by empirical work if we are to know what we are really studying—what the educational environment really is. For example, the introduction into a school of automated instructional systems with individual study carrels influences the social as well as the instructional dimensions of the environment, although normally only the instructional dimension has been attended to. We have found, for example, that many

TABLE 4.1
SCRIPTING A MODEL - CONCEPT LEARNING

Task	Required Learner Skill	Involvement Options	Hypothesized Model- relevant Outcome	Content Outcome Boosted
1. Concept Attainment exercise	Discriminate attributes Infer principle Analyze strategy	Reception-Selection Sharing Strategies (Move) Student Leadership	Concept-attribute Concept of classification	Integration
2. Enumerate data	Recall Select by criterion e.g. heterogeneity triviality abstract-concreteness	Individual - Group (Number) Student Leadership	Enumerate data by criterion	Recall
3. Select data	Develop criteria Apply criteria	Group (Number)	Apply criteria	Recall
4. Group data: Initial pass (Includes Label)	Generate Groups	Individual - Group - Class (Number)	Group basis of common attributes	Integration
5. Share - classify - justify - predict labels	Identify common elements of grouped data Generate multidimensional possibilities	Student - talk	Identify principles of classification Concept of classification	Integration
6. Regroup - label data of some group members	Regroup - relabel data	Student - talk	Concept of classification	Integration
7. Multidimensional grouping	Generate multidimensional possibilities	Number	Multiple classification	Integration

8. Share, classify, justify	Identify principles of multidimensional grouping	Student-talk	Concept of multiple class	Integration
9. What is a label? (Share labels)	Analyze labels	Student-talk Leadership	Nature of Label	Integration
10. Alternative Labels (Generate and share)	Generate labels which refer to common attributes Compare labels against criteria of appropriateness	Student-talk	Criteria for adequacy of labels	Integration
11. Group labels	Generate concepts of concepts	Student-talk Number Leadership	Group Abstractions	Integration
12. Share Groups	Analyze concepts of concepts	Student-talk Number Leadership	Criterion for grouping abstractions	Integration
13. Label grouped labels and share	Generate labels	Student-talk Number Leadership	Ability to use labels for grouped labels	Integration
14. Characterization Task (Newspaper Story)	Use labels to operate on data	Student-talk Number Leadership	Application of labels in descriptive tasks	Integration

students do not like to learn in such a setting for purely social reasons.

The closer we get to the study of learning environments, the more differentiated are their elements. In Table 4.1 there appears a task analysis of some of the dimensions of a concept-learning model of teaching.

In the second column are the cognitive tasks which are presented to the learners; in the third column are the involvement options—the options that a teacher has with respect to the social system which will be created. In the first column on the left is an identification of the kinds of learner skills required to respond to the different tasks. This particular sequence of tasks can be carried out in a little over an hour with junior high school aged children. It provides some idea of the complexity of the environment that is created within a simple model used for purely instructional purposes.

In order to validate any given teaching competence, we need to analyze the dimensions of the environment thoroughly and to study them with respect to their effects on the learner. Only in highly controlled studies can we vary small aspects of the dimensions of the environments, but it is necessary to employ highly controlled studies—to conceptualize the environment very clearly and to manipulate variables systematically—if we are to obtain any reasonable knowledge about the effects of teaching on learners.

The Multiple Outcomes of Educational Environments

The learner in a multidimensional environment always produces a variety of kinds of outcomes.⁸ There are substantive outcomes, such as the recall of information and the development of concepts for integrating it; there are attitudinal outcomes, such as a liking for the material, the method, and the teacher. There is the learning of social and cognitive skills. All of these need to be taken into account. Specific educational models may promote particular kinds of learning (what Marsha Weil calls model-relevant outcomes), but many models are likely to share educational outcomes, and may boost the same outcomes about equally. We need to test hypotheses about the existence of model-relevant outcomes and about cross-model effects. It is more likely that particular high-order and affective outcomes will be boosted by specific models than that paired-associate learning will. Exposure to material is likely to be such a powerful determinant of sheer paired associate learning by a highly-motivated learner that it is unlikely that very many models will stand out. The integration of information, however, and the ability to apply information to new situations may vary much more.

I believe that the transfer of learning should receive the greatest emphasis in the evaluation of instructional strategies and hence, teaching competences should be validated much more by the effect on transfer than by substantive outcomes. Consider, for example, the training of teachers. When we try to evaluate the effectiveness of strategies for training teachers, *all* of the important criteria are ultimately bound up in problems of transfer. If we teach an instructional strategy to a teacher, our only way of evaluating his performance is to learn what effects it has on children as the teacher transfers it to new situations.

It is equally important to know what children transfer from any given instructional encounter. How do they carry the information to a new situation; how do they use the concepts they have learned; how do they apply the principles that are being developed; how do the attitudes they have formed and the social relations they have developed affect their subsequent behavior? Simply to evaluate the performance of a model of teaching, or a given teacher who uses it in terms of what is obtained from the encounter itself is to concentrate on the trivial and ignore the important. Probably the most important learning outcomes of models of teaching are those which are embedded in the patterns of the models themselves and affect responsiveness to future situations. For example, one outcome of a counseling environment is to equip the learner to respond to nondirective environments. In the case of learning strategies centered on the modes of inquiry of the disciplines, the most important outcomes are those which have to do with his application of those modes to new situations. The model of learning embedded in the model of teaching is the critical transferable content. In the studies Dave Hunt and I have done recently, we have studied the process by which children acquire a model of teaching to the point where they can actively carry it out by themselves. We have exposed learners with different personality structures to training programs until they are able to behave in the mode of the model independently, without the assistance of the teacher. We have been very much surprised by the relative quickness with which learners pick up the heuristics of a model. What has not surprised us so much, but which complicates our lives considerably, is the difference in transfer by children of different personality types. The more rigid children, for example, tend to apply a model more or less as the teacher uses it, following it almost as a formula, where the more complex children tend to transform a model and apply its heuristics rather than the formula of patterns that they have been exposed to.

The study of transfer is enormously complicated by the fact that the learner transforms what he learns as he applies it in new situations. This plagues the study of transfer in teacher education. A teaching skill which stands out clearly in a specific type of microteaching laboratory exercise may appear in quite a different form and even be almost buried in the events of the classroom. For example, several of my associates and I for several years studied and trained teachers' sensitivity to learners in simulated teaching situations. In the laboratory settings the teacher was given a teaching task and presented with one or more learners (actually role players). The role players emitted cues which represented a point of view which conflicted with the teaching task. Until the teacher read the cues, attained the concept of the learner, and modified his behavior to build a bridge to the teaching task, he was unsuccessful in the exercise. We found striking individual differences in teachers' ability to "read" the learner and modulate to him, and in the styles which were employed to accomplish this modulation. We also found great individual differences in teachers' ability to acquire the "skills" of sensitivity. When we followed teachers into the classroom, however, we found no reliable way to detect those skills. In the classroom the teacher was trying to read and modulate to many children, sometimes simultaneously. He transformed his skills enormously

in order to relate to the larger group of children. Sometimes we clearly saw versions of the behavior we were after; but in general we were unable to develop an economical and fair system with which to record it, let alone to judge its effects on children.

Proposed Focus on Learning Outcomes. Presently, my associates and I are exploring the following types of learning outcomes as we explore the effects of different models.

The substance of the transactions. We explore the acquisition of information (paired associate learning), concepts, and principles that integrate the material which is explored by the learner during any given period of time. Multiple-choice tests are used at the paired-associate learning level and essays or interviews are used to collect data which is content-analyzed to determine the kinds of concepts and themes which the learner develops in the course of his inquiry.

Model-relevant concepts or skills. We have been exploring the development of tests to identify the particular skills and concepts which are generic to particular kinds of models. The synnoetics model of teaching, for example, attempts to teach children to think metaphorically. If they explore a given subject matter in metaphoric mode, they will acquire information and build concepts, but they also learn the concept of metaphor and how to use it. This is a model-relevant concept, and the ability to use it is a model-relevant skill. Group investigation requires learners to work together to define objectives and to analyze one another's perceptions in a social situation. The ability to work with others to identify objectives becomes, therefore, a model-relevant skill for group investigation. Concept learning teaches children to enumerate, classify, and label data. The concept of classification, the concept of labeling, and the ability to classify and label are therefore model-relevant outcomes for that model.

Transfer of substance of outcomes. This outcome involves the ability of the learner to apply what he learns to a new situation. If, for example, we teach children how to analyze a culture we would then investigate their ability to analyze a second one, or perhaps compare the second analysis with the first.

Acquisition of the model of learning. I referred to this earlier when I described our attempt to determine what it takes to help a student learn to use the cognitive and social skills of a model of teaching to guide his own inquiry. A clear example is when we use inductive methods with children and then study their ability to carry out inductive inquiry independent of us. Situational tasks are most appropriate for investigation of this kind of outcome.

Attitudes toward self. Inventories can be used to investigate how the learner views himself when interacting with any environment. Inventories can also be used to study the following affective outcomes:

Attitudes toward material. (How does the student like the substance and learning material he encounters in an environment?)

Attitudes toward the method. (How does he feel about the methods of teaching he experiences?)

Attitudes toward the teacher. (How does he feel about the training agent?)

These are depicted in Table Two.

TABLE 4.2

Type of Outcome	Type of Measuring Device
Recall of Material	Tests - True - False and Multiple-Choice Items
Integration of Material	Content Analyses of Descriptions Given by the Learner
Transfer of Concepts Learned and Social and Cognitive Skills of the Model	Situational Tasks
Attitudes about self	Inventories
teacher	"
material	"
method	"

The Generally-Effective Teaching Skill

What is the likelihood that we will identify a half-dozen skills which have powerful effects across models that will increase learning in many kinds of environments?

As we make a task analysis of particular models of teaching, we find varying degrees of overlap. There are some models which appear to have almost no overlap with others. The skills of some fairly simple models seem to be entirely subsumed by fairly complex models. It is difficult, however, to identify any given skill except in the matrix of the pattern of a model. For instance, negotiating with a learner in a counseling situation is rather different than negotiating with the learner in a group dynamics model while the group is trying to analyze a situation and develop hypotheses about it. There is negotiation involved in both cases and, at a microanalytic level, we can detect similarities. However, a teacher brings that skill into play differently within the different patterns. Marsha Weil conducted an investigation which bears on this issue.⁹ After an analysis of the interaction analysis patterns of a group of teachers who had learned several models of teaching, she extrapolated what appeared to be three general and widely useful skills that were related to difficulties in implementing the models. Each of these, moreover, was a fairly complex skill. For example, the skill of structuring consisted of the ability to implement several types of social relations in a classroom from a condition in which the teacher directed all of the activities to a mode in which the activities were negotiated with the children and to one in which the learners negotiated the activities themselves. The skills, in other words, are not simple skills such as the ability to ask a particular kind of question or probe in a particular way. Weil built

instructional systems which could be used to train teachers to use those skills and did so, rating the degrees of performance that resulted. The teachers then learned several models of teaching and she correlated the degree of skill with the ability to acquire the models. She found that the skills, rather than having general effects as she had supposed they might, were model-specific, then were associated with specific models. We formed the generalization that the more classroom-unusual the behavior required by a model was, the more the skill that was relevant to that behavior was important to the acquisition of the model. In other words, a state of constructive structuring is relatively rare in most of today's classrooms, and such structuring was extremely important in the group-investigation, but much less important in a model that utilized more or less conventional structuring. The skills, in other words, turned out to be model-specific.

Eventually, we probably will find specific characteristics of teachers or types of skills that are quite general. To search for them at this time, however, before we have identified the skills most appropriate to particular models and determined how those degrees of skills relate to learner outcomes, seems premature.

Duration: Avoiding the Quick and Dirty

How long does a learner need to be exposed to a particular environment before we can expect to see changes in his behavior? This question has critical importance for the design of investigations to determine the effects of various models of teaching. To try to validate a competence in studies which are too brief for the effect of the competence to be detected is wasteful. It also tells us little if we employ studies so long that almost any reasonable approach has the kind of effect we are studying. There are, in the literature, an enormous number of laboratory experiments in which the duration of teaching was only 30 minutes or less. It is hard to imagine very many learning outcomes which would be seriously affected by interventions that short.

It should not be difficult, however, for us to engineer sets of studies designed specifically to develop principles of duration. At first these studies should probably be outcome-specific. For example, recall of information contained in written passages might be a dependent variable of experiments in which teachers interact with students for varying lengths of time. This type of experiment should be repeated for varieties of learning outcomes and approaches to teaching. Without the principles which these studies would generate, we really have only rule of thumb to decide how long experiments on teaching should be. If, as it turns out, transactions with the teacher should affect recall by the learner in a half-hour or an hour, we can productively conduct experiments in which the teaching episodes are of about that duration. If, however, it should turn out that one must teach for 10 or 12 hours before the learner learns more than he does by reading the material by himself, then those experiments would have to be much longer.

It would also be important for us to track both the induced effects of teaching transactions (nurtured) as well as those that are directly taught. For example, one nurtures rather than instructs the development of self-

esteem through praise and support, but he can directly instruct the commutative principle. It is reasonable to suppose that nurturant effects occur more slowly than instructional effects. Personally, I think this is why there have been so many unsatisfactory experiments in the field of reading. One does not simply instruct in phonics principles (direct effects). He provides a setting in which children can teach themselves a wide variety of things, including developing their vocabulary, forming concepts about coding and decoding, and developing the kinds of attitudes and reading habits which lead to real reading power. An experiment to evaluate a teaching strategy in the field of reading that lasts only 6 or 8 weeks, one may find very few of the important nurturant effects.

A Systematic Program of Research to Identify Teaching Competence

I propose a six-stage model for identifying and validating teacher competencies. The model is built directly on a resolution of the issues in ways which appeal to me.

Defining the Roles of the Teacher

The process begins with the partitioning of the teacher's role. Ideally, the role is defined as a result of a construction of a model of the school (see Appendix A), so that we are looking for competencies for a characteristic kind of educational institution which deploys its staff deliberately and rationally. However, if we are faced with the problem of searching for competencies within the present general teacher roles, then these need to be partitioned into subroles such as the teacher as individualizer of instruction, the teacher as counselor, the teacher as instructor of groups, the teacher as lecturer, and so on. For purposes of illustration, let us work with three roles: the teacher as an individualizer of instruction, the teacher as a small group instructional leader, and the teacher as a lecturer.

The Specification of Role Outcomes

The definition of the roles should include a determination of the outcomes of the anticipated outcomes of the role: that is, how it is expected to function for the education of the student. For example, the role as individualizer might be associated with the teaching of basic skills in reading and mathematics at the elementary school level and the basic skills associated with any given subject matter at the secondary level. (For example, grammar in English, map skills in geography, and so on.) These role outcomes are extremely important because they will be the basis on which we ultimately wish to validate the competencies which are eventually identified.

Developing Appropriate Settings and Support Systems

The next task is to develop settings in which the role can legitimately be played and to specify the kinds of support systems that would be available to the teacher. For example, a teacher who is playing the role of an individual of instruction and has available to him the resources of the IPI System is

in a very different position than the teacher who has available to him a library of trade books. In both cases the role can be played but the competencies which are needed are likely to be very different. The IPI System provides the teacher with a diagnostic and prescriptive system that he can use on many of his tasks, whereas if he has open-ended resources, he has to determine the skills he will teach and perhaps to build his own diagnostic and tracking system.

Identification of Appropriate Models of Teaching

The next task is to identify which models of teaching enable the teacher to fulfill the role. There are two ways of doing this. One is by studying educational theories and operationalizing models of teaching from them. Marsha Weil and I have made a survey of available models, and we reach into this storehouse and match the existing models with the roles of the teacher.¹⁰ For instance, we are able to identify a number of interactive and information processing models that seem appropriate to the role of small group instructional leader. Concept learning, group investigation, concept attainment, etc., are all appropriate to the role.

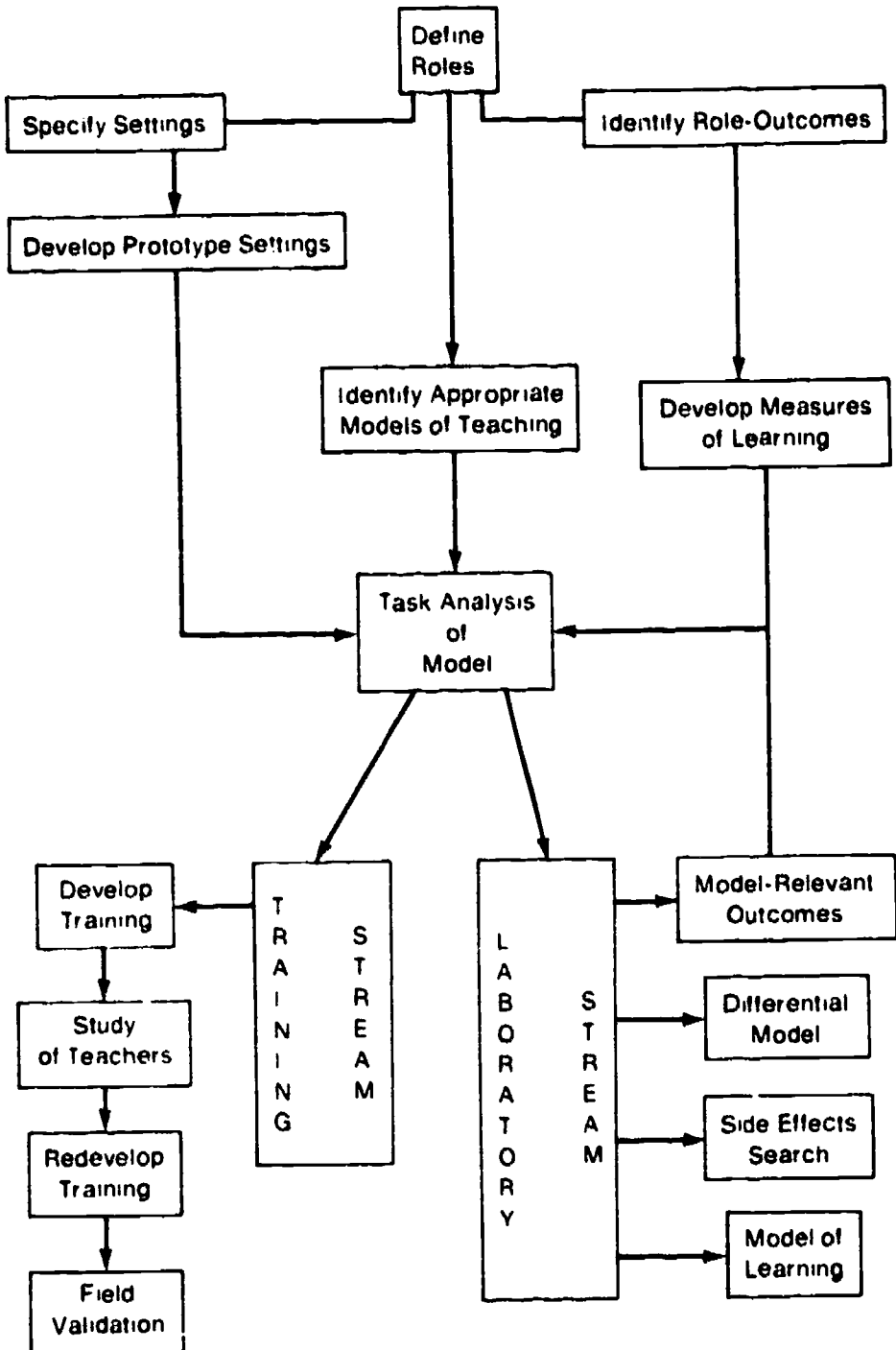
The alternative to theoretical sources is the study of practitioners. We can ask teacher to play the roles in our prototype instructional settings and to identify the patterns of behavior they use, thus developing models originated by practitioners. This is the method preferred by Arno Bellack and his associates at the Center for the Study of Teaching at Teachers College. I am sympathetic to the model, for I am also a member of that Center, but in my own work I tend to use the former approach. Either way, a series of models is generated which are role-appropriate. In the next step, a behavioral analysis of the models is made by comingling two methods. One method is to obtain behavioral examples of expert teachers trying the model in the prototype settings. A group of experts familiarize themselves with the model and try it out, so that we can get some idea what the specific behaviors are that they use. The second method is to make a careful conceptual task analysis of the model, thinking it through stage by stage, and identifying the cognitive and social tasks appropriate to the model.

From these two processes we get a behavioral representation of each model. In the case of some models, it may have a relatively fixed sequence of phases, whereas others are very flexible. If the conceptual task analysis is used, it should be followed by the empirical representation of the model by several expert teachers, so that a range of skills appears as well as a normative pattern.

The Laboratory and Training Streams

From this point there are two sequences of investigation: laboratory and training streams. One sequence is research on the model of teaching. It includes an analysis of the function of different teaching skills in the model and an analysis of the effects of the model on learners in the short and long run. The second stream of work is the training stream, which focuses on the development and testing of systems to train teachers to the model of teach-

Figure 4.1
A PROGRAM OF RESEARCH



ing and the study of the effects when a range of teachers use the model in field settings. The two streams of work need not be separated in practice although it is often convenient to have them separated. Whether they are done together or correlated, they have conceptually different purposes. In the laboratory research stream, studies are made of the effects of the model in controlled settings with an attempt to identify the kinds of skills needed by learners to respond to the model, the differential responses of learners, and the kinds of outcomes which are boosted by the model in comparison with others. Thus the outcomes of the model of research are model-relevant outcomes, a differential model or model for applying the model with children of different characteristics, an analysis of the side effects of the model, and identification of the model of learning. In the training stream the results of the empirical and conceptual task analysis of the model are used to develop training systems to expose both teacher candidates and experienced teachers to the model. Teachers study the model, and they are studies in microteaching and field situations as they try them out. An analysis of their behavior leads to a more precise identification of the skills which they have difficulty learning and which are essential to carrying out the model. The results of this research lead to a redevelopment of the training system. Another trial with the training systems determines the effect of the additional skill training and the effects of the model on learners. To compare the results of the behavior of different teachers, controls have to be placed on the content which is taught, or it becomes impossible to compare outcomes. This phase of the investigation, however, represents the field validation studies. It may well be that, in the field, outcomes are legitimately different than those in the laboratory settings of the model of research stream, but hopefully there is relatively great congruence between them. A comparison of the patterns of teaching in the field validation studies should indicate whether certain teachers possess skills which result in greater learning than do other skills, whether there are certain microskills which lead to greater effectiveness than any given model. As this program of work is repeated with several models, one edges toward a position where he can determine whether there are any cross-model teaching skills of importance.

FOOTNOTES

¹ John Dewey, *Experience and Education*, (New York: Scribners, 1937)

² N. L. Gage, *Teacher Effectiveness and Teacher Education*, (Palo Alto: Pacific Books, 1970)

³ This school is described in Bruce Joyce, *The Magic Lantern* in David Olson, editor *Media and Education*, 1974 Yearbook of the National Society for the Study of Education (Chicago: University of Chicago Press, 1974)

⁴ Michael Apple, 'Behaviorism and Conservatism' in Bruce Joyce and Marsha Weil, eds *Perspectives on Reform in Teacher Education* (Englewood Cliffs: Prentice-Hall, 1972)

⁵ Arno Bellack et al *The Language of the Classroom*, (New York: Teachers College Press, 1966)

⁶ Rhoda Waid, *The Effects of Models of Teaching as a Program for Preparing Teachers* Doctoral thesis, Columbia University, 1972

⁷ David Hunt, Bruce Joyce, Joyce Noy, Joanne Greenwood, Roma Reid, and Marsha Weil 'Student Conceptual Level and Models of Teaching' presented to the 1974 Annual Meeting of the American Educational Research Association

* Rodney Skager. Evaluating Educational Alternatives. Unpublished paper. School of Education, UCLA. 1974

* Marsha Weil. Deriving Teaching Skills from Models of Teaching. A paper presented to the 1973 Annual Meeting of the American Educational Research Association

* Bruce Joyce and Marsha Weil. *Models of Teaching*. (Englewood Cliffs: Prentice-Hall, 1972)

Additional References

A number of recent papers on related issues have been invaluable sources of ideas for this paper

Del Schalock, ed. A Program of Research on the Efficacy and Design of Competency-Based Teacher Education. Unpublished paper. Teaching Research, Monmouth, Oregon. 1974

Rodney Skager. Evaluating Educational Alternatives. Unpublished paper. School of Education, UCLA. 1974

5 Task Force on Relating Teacher Competence to Pupil Outcomes

**Donald M. Medley
David Krathwohl**

A substantial portion of the time allotted to the task force was spent in small groups whose mandate was to formulate questions for discussion when the entire task force reassembled. Since not all of these questions were dealt with by the larger group, a summary is appended to this synthesis.

The issue that seemed to receive the most attention had to do with whether it is defensible to attempt to define teacher competence in terms of pupil outcomes. The wording of the charge given to the task force—to deal with the problem of relating teacher competence to pupil outcomes—implies that teacher competence is (1) a different entity from pupil outcomes, and (2) that the relationship between the two is an empirical question. While the point was disputed, most of the group concurred in the opinion that the teacher should be held accountable for competence rather than outcomes.

An analogy often cited was the medical one: it was pointed out that when a patient dies, the physician is not held responsible so long as he has correctly used the best known methods for diagnosing and treating the patient. It is generally realized that the state of medical research is not sufficiently advanced that sure cures exist for many diseases. Failures tend to be blamed on incomplete research data rather than on the incompetence of the individual physician. When the practitioner's competence is called into question, his defense is based on his knowledge of and ability to apply the best known methods, and not on outcomes.

Some question was raised about the fairness of the evaluation procedures used in the program described by Del Schalock, in which students are re-

quired to prove their ability to produce pupil learning before they may be certified as competent. As one member phrased it, "if the student uses the correct procedure according to his supervisor, and the pupils fail to learn-- who should be held accountable for the failure--the student or the supervisor? Or to paraphrase another comment: is it ethical to demand that a teacher demonstrate more knowledge of the teaching process than those who purport to be training him possess?"

The state of research in teacher effectiveness--that is, research relating teacher behavior (process) to pupil outcomes (product)--is far less advanced, everyone agreed--than research in medical science--that is, research relating the behavior of physicians to patient recovery.

The defense of the Schollock program was that the learning outcomes to be produced were determined by negotiation between teacher and supervisor, and were restricted to what would be reasonable to achieve with the group of pupils in question. This is an important point, and goes a long way toward removing the objections cited above by ensuring (to a degree) that the student is not confronted with an unreasonable task.

Serious questions were raised about the validity and reliability of such a procedure. The reliability is, of course, unknown, and inaccessible to study since each performance is based on different criteria as well as on different pupils, content, etc.; further, not only the evaluation but the definition of the criteria are highly subjective. Decisions as important as whether or not a student is to be admitted to the practice of a profession should be based on measures of high objectivity and reliability, as well as validity.

Questions regarding validity took two forms. The desirability of certifying teachers on the basis of ability to produce only short-term gains in limited content areas and with only one type of class was debated. The fact that the student is required only to produce what his supervisor considers a reasonable amount of learning was seen as tending to stifle progress by requiring the new teacher to be no more effective than the old.

All of this discussion seemed to serve the purpose of sharpening a distinction between the activities of research and evaluation. When we are evaluating teachers individually, we need much greater reliability (and practicality) than we need when we are doing research in teacher effectiveness.

The objective of research is to identify teacher competencies that are related to pupil outcomes in a probabilistic sense, that is, teacher behaviors which are most likely to produce pupil learning. Research studies are statistical in nature and deal with groups of teachers; if we can show, for example, that the mean gain in a group which behaves in one way is greater than the mean gain in a group which behaves in another way, we have evidence of the greater effectiveness of one way of behaving. Reliable mean differences can (fortunately for us) be demonstrated with measures whose reliability for discriminating between individuals is low.

Measures of pupil outcomes are highly unstable; the principle source of errors of measurement are differences between pupils used one time and those used another time with the same teacher. If the average correlation between gains measures based on different groups taught by the same teacher is around .3 (as Rosenshine reports), then it would be necessary to

average gains in seven groups to get a measure whose reliability is .75. Assuming that we are interested in gains achieved over a one-week interval, this means it takes some seven weeks to measure teacher effectiveness. If we are interested in gains over one year, it takes seven years! The impracticality of using mean pupil gains as a basis for teacher *evaluation*—unless one is willing to use very short term gains indeed—is obvious. And yet just such measures can be used in *research* studies to identify competencies the possession of which will increase a teacher's effectiveness.

If a teacher is evaluated according to the competencies he possesses, adequate reliability can be achieved with a reasonable expenditure of time and effort. And when research begins to identify new ways of being effective, we can begin to certify teachers who are more competent—and more effective—than the ones we certify today.

When planning a study in this area, it is important to design a study with sufficient power to detect relationships that are none too robust; in particular, sample sizes of at least 50 to 100 teachers are needed. Data should be collected that provide information, not only on the teacher behaviors and the pupil outcomes which are the primary concern, but also about variables likely to interact with them. Pupil characteristics (especially socioeconomic status and related characteristics) and type of curriculum seem particularly important. It is dangerous to assume linearity. Recent research strongly indicates that in many cases, the regression of pupil growth on teacher behavior is curvilinear with optimum effectiveness achieved by teachers whose behavior is somewhere in the middle of the observed range of behavior. Outcome measures should reflect more than one type of pupil learning, since it is likely that different patterns of teacher behavior will be most effective in facilitating different kinds of learning.

To sum up, the general attitude of the group was positive—the consensus was that research in teacher effectiveness can increase our ability to train competent teachers in performance-based programs by identifying the nature of teacher competence. How rapidly these programs improve depends to a great extent on how extensive a research program we are able to mount and sustain.

Relating Teacher Competence to Pupil Outcomes Task Force

- A. An area of concern to almost all members of the group had to do with the specification of the outcomes to which competencies are to be related. Questions raised include the following:
 1. What philosophical base should be used in determining the desired outcomes?
 2. How may parents (and other community groups) gain access to the establishment of desired pupil outcomes?

3. How will political constraints and decisions based on them—which may or may not be compatible with decisions about desired outcomes—be dealt with?
 4. Are there some competencies which relate positively to some outcomes and negatively to others?
 5. Should long-term outcomes be used; and if so, for what purposes?
 6. How can the public be educated to expand its concept of outcomes to include affective dimensions?
-

B. A second area of concern had to do with interactions between teacher and learner characteristics (and other extraneous variables) and their effect on relationships between competence and outcomes.

1. What variables need to be taken into account in identifying pupil learning styles, teacher learning styles, etc.?
 2. Does the interaction between these two vary for such types of outcomes as self-concept, student attitudes, and mastery of cognitive skills?
 3. Are there any generic teaching skills whose effects are invariant across such variables as these?
-

C. A number of researchable questions were raised, including the following:

1. What kinds of teacher competencies produce what kinds of outcomes with what types of pupils?
2. How are these relationships affected by entering characteristics of pupils and curriculum variables, and how do they interact with teacher behaviors?
3. Is it possible to isolate the effects of teacher behavior from other environmental factors?
4. What are some of these other variables?

5. Is there enough known correlation between competencies and outcomes to justify certification on the basis of competencies?
 6. Which kinds of outcomes are most reliably measured at the pre-service level? at the inservice level?
 7. Are there competencies which once acquired are never lost? Are there competencies which need continual reinforcement and practice to be maintained?
-

D. Miscellaneous Questions

1. How can effects of setting and teacher role be separated?
2. What variables are involved in research relating teaching and learning styles?
3. Should the emphasis be on relating outcomes to competencies as such, or in identifying effective teachers and determining what competencies they possess?

Appendix

Summary of Questions Raised by Task Force A

1. Using the regression model, couldn't you regress the teacher behaviors on performance of the outliers?
2. How to overcome the resistance to judge certification eligibility based on pupil learning—uncontrollable conditions, differences in learners, home background.
3. The "invasion of privacy" issue which will emerge when we use data of practicing teachers: effect on career, certification, salary, unless all in population are included.
4. Are there ways of relating change in pupil performance to causal factors without experimental designs?
5. What about styles of teaching vs. child learning styles?
6. What are the act-categories under consideration?
7. What concepts of curricular structure and instructional structure are being assumed?

BEST COPY AVAILABLE

8. Is the teacher assumed to be performing in his individual acts under certain credible principles and ethical norms?

6 A Practical Approach to a Complex Problem

Part 1

George Gustafson

A teacher training program must be considered from three points of view: program development, program research, and program accreditation. California is unique among the states in that it has legislation which is both specific and, at the same time, provides opportunities for modifications in the three areas of development, research, and accreditation. One of the initiating elements for this legislation, the research and development of competency based teacher education, allowed us to move in some very constructive directions in a state where there are approximately 75 teacher training institutions. These institutions range from the very small to some of the major universities—the University of California, Stanford, and others.

The teacher certification legislation in California established the California Commission for Teacher Preparation and Licensing which is responsible for the certification of teachers. What the Commission tried to do was to design a program cycle that would allow a whole new series of events to occur in terms of development of professional education activities. This program cycle contains four pieces in a simple paradigm. (Figure 6.1).

The first of these pieces is planning, during which the institution operates from a set of guidelines developed for, and adopted by, the Commission. These guidelines provide direction to the program planning and development process. During this stage the staff of the Commission works with the institutions to develop a plan for teacher education. The institutions then submit the plans to the Commission for approval, for conditional approval,

PROGRAM CYCLE

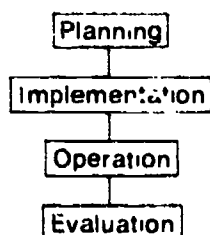


Figure 6.1

or for modification, based on the program planning guidelines. Program planning may take as long as two to three years.

The California legislation has a very interesting and thoughtful part to it, the requirement for involvement by persons outside the traditional teacher education program. This requirement is incorporated into two aspects of the guidelines for program development by teacher education institutions. The first guideline requirement is that the world of practice must be involved in the teacher education program. The teachers, school districts, and the community must be involved in the planning process. The people in the field have some impact upon the development of the teacher education program by providing the institution with information about the types of teachers needed in the schools and the training requirements for such teachers.

The second guideline requirement, which I think is equally important to the university and college people, is that teacher education must be a total campus commitment. The teacher education program may not be developed in isolation but must result from a coordinated, campus-wide, participation in the developmental process.

The planning stage then moves into the second stage, implementation, during which the Commission staff works with the institutions in implementing and modifying the approved plans. As modifications become necessary during this implementation stage, such modifications are in turn planned and implemented.

The third stage in the process is that the institution does in fact have an operating teacher education program, based on the plan and the modifications made during implementation of the plan. In the traditional sense, we said that the program begins with a series of ideas, plans, and operational notions, which, as they begin to be implemented, cause courses which are presently being offered to be changed and moved into new directions.

The real trick is to be sure that the programs which are developed, implemented, and operating at the universities have some relationship to the field. To insure the relationship of the programs to the field of public school teaching and to provide for program evaluation, we built in a system of collecting data which involves several steps—the collection of data about the person, the process, and the product

DATA COLLECTION SYSTEM

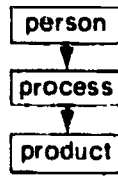


Figure 6.2

The first element to be considered in developing a program for teacher education deals with the person, so you collect data and information about the people with whom the program of teacher education will be working.

Then you do an analysis of the process. The colleges and universities begin to do things to the persons who are in their program. These include some good and some bad, but they are doing these to the students in varying degrees. We have had some interesting battles over the process, and over what the institutions plan to do with the students in their programs; student teaching is one example. Everyone talks about competencies and the need to define the competencies which students must master, but the battle over student teaching centered on whether or not student teaching must consist of 15 units. The law clearly states that the requirement for student teaching shall be one semester of full-time student teaching or its equivalent. But the law does not define the equivalency, and the battle raged over whether it should be 18 quarter units or 12 semester units or what. There was little concern for the planning for the activities in which students would be involved during the student teaching experience.

The third stage is the product which is the outcome of the person going through the process. One requirement of the program development guidelines is that the institutions must state what their students will be able to do when they complete the program, the exit level skills or competencies. There is no attempt at this time, and hopefully there won't be for a long, long time, to specify the competencies which students must attain when they leave a program. There is no hidden list that specifies that these are the 25 or 30 important competencies. The Commission is really working from the commitment that the institution must identify the types of competencies or skills which they presume to be useful and viable for teachers. The person who enters the teacher education program and the process of the teacher education program combine to produce an individual who will teach in a school in California.

The Commission will be concerned with collecting data to provide information about how graduates of teacher education programs are functioning within the public schools of the state. Thus, by looking at the person entering the program, the process of teacher education, the exit skills of the person leaving the program, and the ability of the teacher to teach within the public schools, the pieces fit.

We have a system which has some inherent possibilities for flexibility. Ultimately, the Commission will have to look at the success of the process of teacher education as described by the institutional plans and as imple-

mented within the teacher education institutions. They will have to look at the person, the process, and the product, in terms of how well the graduates deal with the real world of teaching in public schools and how well they can teach. The information obtained by looking at the product, the teacher in the public schools, will provide for a recycling of the teacher education program.

The need, then, is to design a system, a manner of collecting information, which will put all of the pieces together, which will look at the characteristics of the entering teacher education students, the process of teacher education, the skills or competencies of the graduating teacher education students, and the success of the teacher in the public schools.

As one step to accomplish this enormous task of systematizing the data collection process, I have designed a matrix. (Figure 6.3) There are seven categories in which we collect data about what is occurring within the teaching process. The first is one that is fairly traditional, that teachers are individuals and as individuals, they have certain kinds of knowledge, whether it is academic knowledge, public school required knowledge, or knowledge of the anticipated student learning in terms of the instructional objectives of the school or district. Teachers must have a knowledge base from which to organize their teaching, both academic knowledge to convey to students and knowledge of the student learning outcomes which are important with the particular school district.

A second category is knowledge about learning. If teachers are going to teach, they must have some knowledge about the learning process. They must know the ways in which a teacher can assist with the learning process. They must know what stimulates and what depresses learning.

The third category deals with student motivation. If teachers are going to be effective, they must have some operational control over the factors which affect student motivation to learn.

The fourth large area of concern is one about which there is a lot of talk today, diagnosis. The talk centers on the need for professionals to be skilled at diagnosis, and the need for the development of tools to aid in the diagnostic process. Teachers need to identify the level at which the student is functioning in a given subject or skill area.

I think it is almost impossible to deal with the business of diagnosis without looking at the same time at prescription, category five. One of the most frustrating things for me as a professional, particularly as I consider the process of individualization of instruction, is finding that we have a large number of diverse diagnostic procedures and that we end up with the same basic prescription: the assignment of a page or series of pages in a textbook. If you can't prescribe alternative strategies to meet the diagnosed needs of students, there is no point in doing all of the elaborate diagnosing.

The next category, six, deals with communication. Teachers must communicate with a variety of audiences: other teachers, students, the administration, and, very important in today's world, the community. So we must begin to collect information about the teacher as a communicator.

The last category has to do with evaluation. Teachers need some knowledge of and some skills in evaluation of both individual student and total class progress. Thus, this matrix provides a means whereby we can divide

Figure 6.3

WHAT ARE WE MEASURING?

SUBJECT MATTER (CONTENT)	The Person	College Training	Community Needs	District Requirement	Pre-Employment Training	Post-Employment Training
Knowledge						
Academic						
Public School Required Knowledge						
Public School Objectives						
LEARNING						
Stimulators						
Repressors						
MOTIVATION						
Theories						
Practices						
DIAGNOSIS						
What can he do?						
What does he need? (Short-Term)						
What is possible now? (Interim)						
What is required for him? (Long-Term)						
PRESCRIPTION						
Learning Style						
Learning Materials						
Alternate Ways						

COMMUNICATION

	The Person	College Training	Community Needs	District Requirement	Pre-Employment Training	Post-Employment: Training
Learner / Teacher						
Learner / Learner						
Teacher / Teacher						
Teacher / Administrator						
Teacher / Community						

EVALUATION

Student Progress						
Group Progress						
Class Progress						
School Progress						
Testing / Appraisals						

into seven basic elements those teacher skills which institutions are telling us are important teacher skills. Even though each institutional plan is unique, and the experiences through which students progress within the different institutions on their way to becoming teachers are diverse, we have a consistent system for categorizing what institutions tell us are (1) the learning experiences of their students and (2) the exit skills of their graduates. The development of a system to categorize the stated experiences provided within institutional programs of teacher education and the anticipated skills of students upon completion of the programs was tough, given the uniqueness of each institutional plan, until we developed the matrix, which allows for classification of the process steps and the product definitions.

When we look at students as they leave teacher training and enter the real world of teaching in public schools, we must recognize what everyone knows, that districts have requirements and constraints which are unique to the individual districts. When we begin to collect information about the skills of the teacher at the completion of a teacher education program, then we can design a pre-employment training program at the district level to meet the specific, unique, needs of the district in terms of teacher competencies.

The categories in the first matrix define the broad areas of necessary teacher skills, but do not define any of the situations in which teachers

Figure 6.4
INSTRUCTIONAL MATRIX

	Learner Needs	Learning Objectives	Curriculum	Materials	Staff	Space	Time	Costs
I.	EVALUATION							
<i>Standard Instructional Setting</i>								
A. Large Group (15 +)								
B. Small Group (Less than 15)								
C. Independent Study (Individual Work)								
II.								
<i>Library Resource Setting</i>								
A. Research								
B. Books and other flat materials								
C. Programmed instructional material								
III.								
<i>Multi-media Center Setting</i>								
A. Packaged Material								
B. Tapes, Film Strips, Film, etc.								
C. Material Production								
IV.								
<i>Laboratory Setting</i>								
A. Experiment								
B. Test								
C. Manipulative Materials								
D. Concrete Experiences								
V.								
<i>Community Setting</i>								
A. Application of Skills								
B. Need Assessment								
C. Involvement								

teach. To do that, we have designed another matrix. As identified within this matrix, (Figure 6.4) there are five kinds of instructional settings in which teachers are asked to function in public schools. While the matrix was designed with self-contained classrooms in mind, it can also be used to examine open space learning environments and alternative school settings. Whether there are walls or not, we are still dealing with the five instructional settings.

The first is the standard instructional setting where teachers instruct a large group of students, a small group of students, or work with students on a one-to-one basis. In this setting, teachers use a series of instructional methodologies and resources to meet the needs of the students within the instructional setting.

The second setting is the library resource area. In such an area, the student is conducting research with traditional research materials such as reference books, encyclopedias, trade books, and programmed materials. Library resource areas are found in self-contained and in open space instructional settings.

The third one is the multi-media center, distinguished from the library center by the presence of several modes of presentation of materials, such as movie and film strip projectors. While some multi-media centers are stocked with the latest technological advances, others consist of less equipment and the sharing of equipment among teachers or schools. Students may move to an area specified as the center and containing all of the equipment, or the media materials may be brought into the classroom.

The fourth setting, one of the most important to me, is the laboratory where the student has the opportunity to experiment directly, a place for children to develop skills in stating and testing ideas. This is, again, the type of activity which can be conducted anywhere and does not require a fancy location complete with the latest equipment.

The last category deals with the community. The community is any instructional setting outside of the school. The need to utilize the community as such a setting increases as we involve the community in defining the student needs. These are the needs which the instructional program must meet in developing the goals for that program. We cannot involve the community in establishing the needs and goals of an instructional program and then attempt to isolate the program from the community, or educate the students without concern for an involvement in the community. Again, teachers in traditional school organizations, as well as teachers in open space programs, can make use of the community for a variety of instructional activities.

These two matrices define the categories of teacher competencies and the instructional settings in which those competencies are employed. Jointly, they allow for the classification of the information obtained from the program plans from teacher training institutions, the data about program implementation, and the information about the success of graduates when they are teaching within the public schools. This is the first step in a system to obtain and organize information about the totality of teacher education, looking at the person who enters the program, the process of teacher education, and the product, the teacher in the public school.

Part 2

Marjorie Powell

Teacher training institutions admit students who meet entry criteria and then move the students through the process, their teacher education program designed to prepare students to teach in the public schools. The California Commission for Teacher Preparation and Licensing requires that institutions describe the program which they will implement. The Commission grants approval for a period of time to the program as described by the institution, and is establishing a procedure to insure that programs are implemented as described, or that modifications are documented.

While these steps involve the person entering the program and the process by which the person is educated, they do not yet look at the product, the teacher teaching within the school classroom. The important question is still unanswered. Will the students who move through the program as described and implemented by the institution be successful teachers when they have received their certificates from the Commission and have been hired to teach? What skills are necessary to be a successful teacher?

In order to look for the answer to the question of what is a successful teacher, the Commission has undertaken a research study, the Beginning Teacher Evaluation Study, with funding from the National Institute of Education. The purpose of the multi-year study is to identify teacher behaviors which are related to student achievement in reading and mathematics.

We have limited the focus of the study to reading and mathematics because of the necessity of confining the research to manageable proportions. Such a study has a way of expanding until the entire universe is included and the impact of the study is diffused to the point where nothing is learned. However, we do recognize that the elementary school curriculum contains more than the basic skills of reading and mathematics, and we hope that the work of this study will be expanded in the future to incorporate more areas of the elementary program.

The first year of the multi-year study, Phase I, was devoted to planning. After the development of an initial design for the study, the Commission sponsored a Research Design Conference in February, 1973. The conference participants, educational researchers, teacher educators, school administrators, and subject matter specialists, reviewed the research design and made a series of recommendations. Based on many of these recommendations, a design for the second year of the study was finalized.

The second year, Phase II of the study, focuses upon the development of instruments and the generation of research hypotheses. Educational Testing Service (ETS) was selected to conduct the work of Phase II during the 1973-1974 school year. Experienced teachers in grades 2 and 5 form the sample for Phase II, the first year of field work. School districts were asked to participate, then experienced teachers were contacted, provided with information about the study, and asked to volunteer.

During the planning for Phase II, a variety of procedures was discussed for identifying the teachers to be included in the sample. We considered a

process of nomination by all of the teachers in the school or by the principal. However, even when the criteria to be used to make the nominations are clearly specified, in many instances nominations are based on factors other than the specified criteria. We finally concluded that we wanted to study experienced teachers in the variety of instructional situations, without using nominations to identify teachers to be included in the study.

The Phase II sample of 97 experienced teachers in eight districts included teachers in urban, suburban, and rural districts teaching students from low and middle income areas. Some teachers are in self-contained classrooms while others teach in open space programs. Some have multi-grade groups of students, some are departmentalized, teaching only one or two subject areas. Several teams of teachers volunteered and are included in the work this year. The majority of the teachers have some form of teaching assistance—an aide, parent volunteers, or older students who tutor specific children.

We are particularly pleased that teachers from so many instructional settings are involved in the sample during the first year of field work. We hope to identify teacher behaviors which are important in a variety of instructional settings and must insure that the measurement instruments which we use to identify and measure teacher behaviors are appropriate for, and applicable to, a variety of instructional settings.

While the major focus of the study is on newly certificated teachers, a model must first be built, based on behaviors exhibited by, and related to student achievement in the classrooms of experienced teachers. For this reason, experienced teachers will be studied during Phase II and will be the major part of the sample in the work next year.

Instrument development during Phase II has included the development of student achievement tests in reading and mathematics, specifically designed to measure student achievement in reading decoding, comprehension, application of skills and attitude; mathematics computational skills, concepts, application of skills, and attitudes.

Fall data collection was conducted during November, 1973. Student achievement will be measured again in the spring to provide a measure of student growth during the school year. Most of the student instruments have been developed specifically for this study, so normative information will not be available.

Instruments have also been developed to measure teacher knowledge of the teaching of reading and mathematics, and teacher verbal ability, or the ability of the teacher to process information. The latter instrument included subtests of vocabulary, memory, classification skills, inductive and deductive reasoning, and creativity. A film test is being developed to assess the skills of teachers to diagnose the reading problems of students and the range of alternatives which the teacher would prescribe to meet the identified need(s) of the students. The test is being developed because we recognized the importance of the skills of diagnosis and prescription, but we were concerned that we might not observe enough instances of diagnosis or prescription to adequately measure the skills of the teachers during the somewhat limited observation period in the classrooms.

Observation procedures and two instruments have been developed by ETS

BEST COPY AVAILABLE

specifically for the multiple classroom observations within this study. The observations are occurring from February through April of the 1973-1974 school year. Differing numbers of observations are being done, using the two observation instruments. The teachers indicated that they would prefer as many observations as possible, since they did not believe that a few observations would provide enough information about important teacher behaviors. The differing observation patterns will also allow for the analysis of the reliability of the instruments and the observers, and a determination of the varying types of information obtained through the use of each instrument when the same classroom is observed at the same time by observers using different instruments.

Teachers are also being videotaped, so that further analysis of teacher behavior can be completed. The teachers will have an opportunity to view their own videotapes after the taping is completed, providing the teachers with an opportunity to see themselves in instructional situations. In addition, information about both teacher and student background has been collected. A series of work diaries has been developed to collect information about the instructional program within the classroom throughout the school year and the particular instructional events during the specified weeks of the year. The teachers work within an instructional setting, with resources available to them, and with specific goals for students. The work diaries are designed to obtain data concerning the instructional resources and goals.

The school climate is being measured through questionnaires for teachers and principals and an interview with the principals. All school instruction occurs within the setting established by the persons involved in the school, the principal, the teachers, and the students. This environment, or climate, has an impact upon the instructional situation and therefore must be assessed in order to adequately determine the relationship between teacher behaviors and student achievement.

The consistency of teacher behavior throughout the school year will be determined through the consistency substudy, which involves the videotaping of a sample of teachers three times during the school year, and the analysis of the videotapes in terms of the consistency of teacher behaviors across the three videotapes.

The stability of teacher performance, as measured by student achievement, will be assessed through the collection of historical test data for students in the classrooms of the sample teachers in previous years. This task is especially difficult due to the differences in testing programs across the eight participating school districts.

The dependent variable in the Beginning Teacher Evaluation Study is student performance, a measure of growth from fall to spring. The independent variables are teacher performance (behaviors), teacher characteristics (years of experience, etc.), student characteristics (age, language background, etc.), and school characteristics (size, special programs, etc.).

Phase III will be designed to refine instruments and to test the research hypotheses which are developed in Phase II, to provide empirical data concerning the relationship between teacher behaviors and student achievement in reading and mathematics. During Phase III, beginning and experi-

BEST COPY AVAILABLE

enced teachers will be included in the sample of teachers, to test the hypotheses with both types of teachers

Phase III is planned to cover three years, with the first year devoted to additional development and field testing of instruments and research hypotheses. A series of special studies are planned for the first year. All of this work will lead into the second year of Phase III, during which a large scale field study will provide a test of the research hypotheses. The third year will focus upon a review of the implications of the study for teacher certification and teacher preparation.

To review these implications and to obtain information about training procedures, a series of small studies will be conducted in which teachers will be trained to utilize the behaviors identified during Phases II and III. These studies will provide both a test of the relationship between the teacher behaviors and student achievement, and information about one or more methods of preparing teachers to exhibit these behaviors.

While I am confident that the study, over several years, will identify important teacher behaviors and will provide information about procedures for training persons to implement the behaviors, I would be unrealistic if I expected that the study would result in a firm list of behaviors which could be defined as vital for all teachers in all teaching situations. Rather, I anticipate that we will identify some behaviors which are important to all teaching situations and other behaviors which are important to specific instructional settings or the teaching of specific skills. I am convinced, further, that institutions will want information about the one or more methods of teaching students to exhibit the identified skills.

7 Classroom Observation

Robert S. Soar

A point of view which has been expressed with increasing frequency is the view that systematic observation of classroom behavior has not been a useful measurement procedure for identifying teacher behavior which is related to pupil growth. Rosenshine¹ concludes that ratings are more effective than observation of teacher behavior using specific behavioral items, and Heath and Nielson evaluate work in this area and conclude,

Last, given the well-documented strong association between student achievement and variables such as socio-economic status and race (see Coleman and others, 1966), the effects of techniques of teaching on achievement (as these variables are conventionally defined) are likely to be inherently trivial.²

Whether this extreme position is accepted or not, it appears that research which has attempted to relate measures of classroom behavior obtained from relatively specific behavioral items to pupil growth may have failed to produce as much as it could have because of a number of methodological difficulties. Since these apparent pitfalls in the past use of observational methods are likely to be related to the success of these methods in PBTE, they will be reviewed briefly here.

Methodological Problems in the Past Use of Systematic Observation

There have been numbers of weaknesses in the studies which have employed systematic observation in the past, including the power of the

typical study, the nature of the observational measure, the nature of the achievement measure used as an outcome, the statistical analyses employed, and others.

Power. A major difficulty with research using classroom observation has been the limited power of the typical study. For example, in the review cited above, Heath and Nielson divide their summary into studies with 15 or more teachers and studies with fewer than 15 teachers. If 15 is taken as the minimum number for a study to be taken seriously, as their division of studies implies, and if the correlation between two variables *as measured* is .30 in the population, then each study will have only 19 chances out of 100 of finding that result significant at the 5 percent level.³ But this is not the worst of the problem. Cohen argues persuasively that if the relation between two *true measures* is .30, a reasonable degree of fallibility in our measurement procedure would reduce the *observed* relation to .10 between the variables *as measured*. If that value is taken as a reasonable estimate of the correlations which we attempt to test, as Cohen argues, then only 6 out of 100 relations which are real would be found to be significant. There is, of course, the countervailing influence that when large numbers of statistical tests are done, some will be significant by chance. Thus, one influence tends to produce too few significant results, and the other tends to produce too many. Since both of these influences are at work simultaneously, it would be surprising if there was very great consistency of results across studies in which the numbers of teachers are small. This, of course, supports the conclusion that our knowledge base is weak; but it argues that the basic difficulty may not be the insensitivity of the measurement tool, but the fact that so many studies have been done which possess limited power (and that these studies are taken as seriously as studies with greater power).

Simplistic Statistical Analysis. Another major weakness of past classroom observation research is probably the statistical analysis, which has usually stopped at the level of linear correlations. Very few analyses have tested the possibility that relationships might be nonlinear, and even fewer have examined interactions—the possibility that the simultaneous effect of two behavior measures might be different from the effect of each taken alone. Intuitively, both ideas seem appealing. It seems unlikely that a teacher behavior which is desirable in some amount is more desirable as the amount increases, without limit. If some praise is good, is more praise always better? is more indirectness always better? is more drill always better? We have found strong nonlinear relations between a teacher behavior which controls or directs pupil behavior, and pupil growth, in four different sets of data.⁴ Yet a linear correlation assumes a “more is better” relationship.

It also seems probable that the meaning of any particular teacher behavior is moderated or influenced by other classroom behaviors. Simple, linear correlations also fail to recognize this possibility. For example, in our recent work, the extent to which the teacher was central in the work of the classroom was unrelated to any measure of pupil change.⁵ But if the teacher was often central and also often chose the problem to be worked on (not necessarily in the same activity), pupil growth was less than if the teacher did one or the other with some frequency. A number of such inter-

actions agreed with each other and with the nonlinear relationships in indicating that greater amounts of pupil growth were associated with intermediate amounts of structure or control by the teacher.

There were also numbers of interactions between pupil characteristics and the classroom behavior which was associated with most gain. For example, increasing amounts of gentle teacher control (a very gentle, unobtrusive, noncoercive style) were more strongly associated with greater subject matter gain for disadvantaged pupils over the summer than for advantaged pupils.

Empirically, the same study found that simple linear correlations produced about one and a half times as many significant results as would be expected by chance, whereas an analysis of nonlinear and interaction terms produced between three and four times as many significant results as would be expected by chance. So it seems likely that another problem in past research has been the infrequency with which analyses have adequately dealt with the complexity of the relationships that exist between measures of classroom phenomena.

Complexity of Achievement Measures. Still another problem which seems likely to bear on the application of observation in PBTE has been the failure to specify the nature of an achievement measure beyond merely naming the subject matter represented. Within a subject matter, a measure may involve learning which is simple and concrete, or complex and abstract, requiring association or information processing. For example, an arithmetic measure might reflect only knowledge of the multiplication table at the level of rote memory, or it might represent complex problem solving. It seems probable that rather different styles of teacher behavior would be associated with most gain for each kind of learning. Three separate studies have supported this idea with varying degrees of completeness.⁶

In the research which has been reviewed in this field, if measures differing in complexity or abstractness were used as outcome measures, it would not be surprising that there is little consistency in the results.⁷

Complexity of Classroom Behavior. Still another possible source of difficulty in past observational research may have been the failure to distinguish between subtly different classroom behaviors. For instance, teacher behavior which was directed at controlling the behavior of pupils, can usefully be distinguished from teacher behavior which was intended to control or limit pupil thought processes. It is possible for the physical activity of pupils to be restricted, and yet for them to be engaged in complex problem solving or in very creative endeavors. In fact, previous data indicated a positive relation between a "freeing" style of teacher-pupil interaction and pupil growth in creativity, but a negative relationship between degree of pupil physical movement and the same creativity growth measure.⁸ There is no way of knowing the frequency with which this distinction is maintained at a theoretical level, but several sets of data indicate that teachers in classrooms rarely make this distinction.

In general, distinctions of this sort which have emerged from the use of classroom observation, seem likely to clarify the nature of the teacher's task and the recording task which represents it. On the other hand, it seems possible that failure to recognize such distinctions may have led to some of

the confusion in results which is present in current literature. Not infrequently, conflicting results are a consequence of the failure to recognize additional important variables, and it seems possible that this may be the case here.

Single versus Multiple Coding. Another major problem which has not always been faced in the use of systematic observation has been recording specific behaviors in isolation. For example, the frequency with which a teacher asks analytic questions may be a useful outcome measure to assess the consequences of a module intended to teach students to ask analytic questions, but perhaps the fact that the student-teacher asks analytic questions will not make very great differences in the achievement gain of pupils, unless these questions are asked at appropriate points in a developmental sequence of learning experiences. For example, Taba and others concluded that a teacher would be unable to maintain a higher level of thinking in pupil inquiry if an insufficient amount of attention had been given to gathering the requisite facts before the higher-level thinking began.⁹

Recent developments in multiple coding (for example, simultaneous, coding who speaks to whom, the level of thinking in the interchange, the affective tone, etc.) are important attempts to record such complex and sequential information and have recently been reviewed by Flanders.¹⁰ The beginning work was the matrix tabulating procedure of the Flanders System which captured, one step at a time, the sequence of occurrences in classrooms.

But there is a possible problem in the nature of the items which are used in instruments intended to record multiple aspects of an event, or sequential events. The problem is that the items which represent each aspect (who and to whom; whether direction, instruction, or question; the affective tone, etc.) must be small in number in order to make the coder's task manageable. This means, in turn, that they must be broad. The fact that a direction was given, for example, would be given in a single code. In contrast, in the *Florida Climate and Control System*, direction giving has 20 codes to scale the degree of coerciveness with which verbal directions are given, from very gentle, noncoercive ones to very harsh, coercive ones.¹¹ Correlations of items with pupil outcome measures in the lower grades suggest that gentle control items relate positively to growth, intermediate items do not relate, and harsh, coercive ones relate negatively. As a consequence, when all direction giving is assigned to one code, relations with outcome measures may be masked by the cancellation of positive and zero, or negative, relationships. If, as we have come to believe, the kinds of observational items which are most likely to be effective in identifying classroom behaviors that will relate to pupil growth will often be items which are relatively narrow, specific, and subtle rather than obvious, then there is likely to be a trade-off in loss of information by the use of items which can be used to multiple code events or sequences. In addition, it seems likely that the multiple coding of events or sequences will complicate the problems of analysis. Each combination becomes a variable, so that combinations must be consolidated in some way, and information may be lost.

It seems clear that two approaches to dealing with the complexity of classroom phenomena and relationships are in use. One records events in

greater detail, but without a multiple record, then develops combinations and interactions by statistical analysis; the other records multiple aspects or sequences of events and then combines them. It seems clear that the nature of the data will differ, and which approach will be more functional, or more functional for what purpose, is an empirical question. Several instruments of both sorts will be reviewed later.

Reliability versus Validity. Traditional concerns over reliability may also lead the developer of observation instruments astray. The items which are most easily observed reliably are the relatively obvious, global, molar ones such as, "Is the teacher front and center in the teaching process?" While observers agree readily on how such an item should be tallied, the measure does not seem likely to be a useful one. Indeed, there are some suggestions that the teacher behaviors which most often relate to pupil gain are the relatively subtle, fleeting ones, in contrast to the more global items for which reliability is easier to obtain.¹² Thus, it seems likely that if the developer of an observation instrument focuses first on the reliability of items as he is developing them, he may risk giving up potentially powerful items and retaining ones which have little utility. Ultimately, in applications in PBTE, the decision should surely be influenced by the question of what is to be measured, but the temptation to give up items which are difficult to observe reliably may need to be resisted.

The Place of Observation in PBTE

Observation may play at least two, and perhaps three, roles in PBTE. First, it seems likely that the observed behavior of the student-teacher in training will often be used as a measure of the success of a module or a larger unit of the program. That is, it would be an outcome measure assessing the success of the student and of the training material. Second, in the long run, it seems likely that observational measures will need to be employed to test whether the differences in teacher behavior which the program produces, in turn, are related to differences in the achievement of desirable objectives by pupils. That is, in the simplest terms, observation may help to answer two questions: does the program make a difference in the behavior of its graduates? and, do these behavioral differences in the graduates make a difference in the pupils they teach?

It seems wise to keep these two functions separate. Each alone will probably be a difficult attainment to assess, but to attempt to relate program differences to differences in pupil outcome seems almost to foredoom the validation attempt at the start. The slippage is great enough at each link in the process to make the demonstration of relationships difficult, but if both sets of linkages are included in the experimental test, the difficulties appear to be increased materially.

Finally, observation systems may be used as a way of teaching desirable behaviors to student teachers in training. There is a growing body of literature indicating that simply teaching a student a set of categories which describe behavior is surprisingly effective in modifying his behavior in a

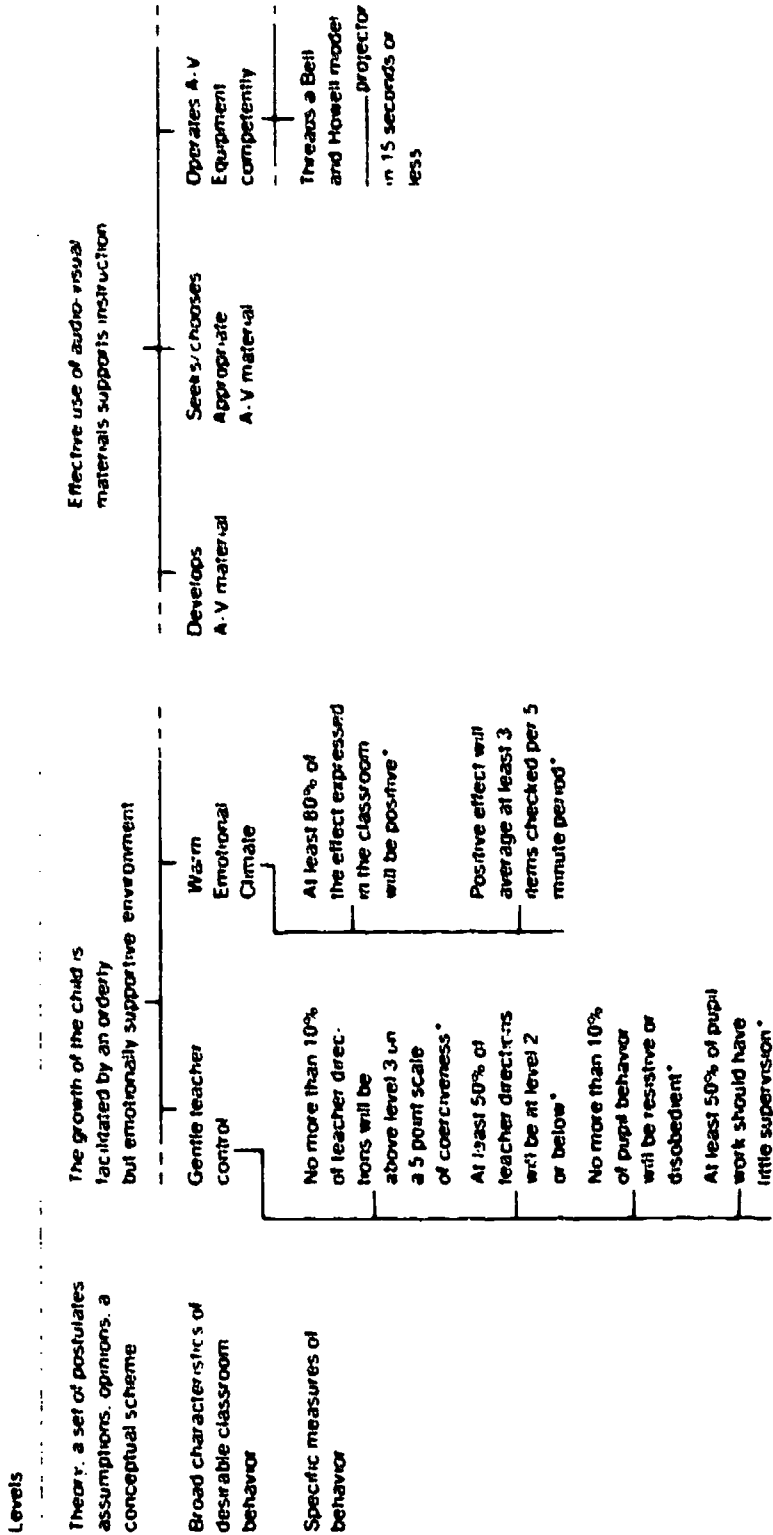
The Usefulness of Hierarchies of Items

As the faculty at our university worked through the question of whether they wished to commit themselves to working on a performance base, the objection was sometimes raised, "I'm not interested in behavior at the level of whether the student can thread the film through the projector." In order to provide a means for talking about this problem, Figure 7.1 was produced. It should not be taken as a set of standards to be taken seriously, but only as an example of a way in which specific behaviors might be organized into successively more abstract and comprehensive conceptual schemes. The top row across the table represents a kind of broad hypothesis or belief about a desirable skill for a trainee to have or a desirable situation for him to create in a classroom. It would be accepted by most educators, but because of its level of abstraction, would probably have different meanings to the different people who accepted it. It is not, as it stands, measurable. If we move down in a row in the figure, however, the concepts of Gentle Teacher Control and of a Warm Climate are somewhat more distinct and behavioral, and yet are sufficiently comprehensive that most people attach values to them. If we move down to a still lower level in the table, a series of measures is identified which makes up each of the desirable classes of classroom behavior. Each of these behaviors, in turn, is defined in a manual which specifies with some detail the settings and the aspects of the behavior which qualify for each of the measures cited. It is a kind of implicit "appendix" to the measurement scheme. Given such a system, the term Gentle Control or Positive Climate has relatively clear meaning, and there is relatively little disagreement about what the term means when it is used. At the same time, it represents behavior at a level of generality that appeals to many people as being of some consequence.

In discussions with others, it appears to be fruitful and to increase understanding and communication to clarify several levels of the hierarchy. If specific measurable behaviors are cited, it helps to give them meaning to answer such questions as, "Why would you care about that?" "Why is it important?" On the other hand, when broader labels are used, greater specificity is sometimes achieved by asking questions like, "What would a teacher be doing if he were doing that?" or "What would he be doing that would be different than if he weren't doing it?" One advantage of such a procedure lies in avoiding the need to deal with hundreds of behavioral items for most purposes. Rather, a considerably smaller number of terms can be used which stand for the larger numbers of behavioral items that make up each complex.

The critical requirement is to verify statistically that the items which make up a composite are positively interrelated. If not, the unrelated or negatively related items may, at the least, muddy the concept, and at the worst, may destroy it operationally. The frequency with which an item empirically differs from others with which it seems logically to belong underlines the importance of this empirical check.

Figure 7.1
SOME POSSIBLE EXAMPLES OF HIERARCHY IN EDUCATIONAL THOUGHT AND PRACTICE



* As measured by the Florida Climate and Control System

Some Widely Used Observation Instruments

The first three of the systems which are described use large numbers of items, but do not code sequences or provide for multiple codes which can later be identified as describing a single event. The third is an extension of the Flanders System of Interaction Analysis, which was the first to begin to represent sequence, and the last three are more recent instruments which code sequence or multiple aspects of an event. All are "general purpose," in that they are not designed to assess a particular competency or skill, or to describe teaching in a particular subject matter. Several are designed for, or have been mostly used in the lower grades.

The Florida Climate and Control System (FLACCS). This is a modification of the South Carolina Observation Record,¹³ which draw heavily on the Hostility-Affection Schedule¹⁴ and the earlier versions of the Observation Schedule and Record (Medley and Mitzel 1958, private communication).

The first section is organized around the concept of direction and control of the classroom by the teacher and the response of pupils, including the teacher's role and the organization of the class. Twenty items of verbal teacher control are scaled into five levels of coerciveness ranging from "Guides, Suggests," to "Orders, Commands" and smaller numbers of nonverbal items are similarly scaled. The response to teacher control is represented by pupil items, such as assumption of responsibility for classroom order, pupil task involvement, obeying, resisting or disobeying, and the amount of socializing. The second section, affect expression, records the eight combinations of teacher versus pupil, verbal versus nonverbal, and positive versus negative affect.

The Teacher Practices Observation Record (TPOR). This is an instrument developed to measure a teacher's practices in relation to John Dewey's Experimentalism, with each item describing a teacher behavior that is widely practiced in schools.¹⁵

It records whether the teacher or pupil is the center of attention, the extent to which pupils are active or passive, and the amount of freedom that pupils are permitted. It describes the nature of the problem as organized around the concerns of pupils or the concerns of the teacher or textbook, as well as identifying the difficulty of study topics. Subject matter is classified as to whether the pupils or the teacher assume primary responsibility for locating it, and whether it is taken from a textbook or a wide range of sources. Whether the teacher evaluates the pupils' work or the pupils engage in self-evaluation is recorded, along with the degree to which the classroom tasks are differentiated for individual pupils.

Florida Taxonomy of Cognitive Behavior (Cog Tax). This instrument is based on *The Taxonomy of Educational Objectives: The Cognitive Domain*,¹⁶ as it was modified and extended by Sanders¹⁷ to assess teacher lesson plans and teaching materials. The Florida group converted Sanders' instrument to one for live observation in the classroom.¹⁸ The levels of the instrument are:

1. Memory - recognize or remember information;
2. Translation - alter the form of the material—figurative to literal, behavioral to verbal—but not to change or evaluate the ideas;

3. Interpretation - identify similarities or differences, relate supporting evidence to a generalization, or carry out a specified operation;
4. Application - bring together and use previously learned material such as deciding what mathematical operation is appropriate to solve a problem.
5. Analysis - applying the rules of logic to the analysis of a problem, or inferring feelings or motives;
6. Synthesis - bringing ideas together, but reorganizing or changing them in such a way as to produce something new;
7. Evaluation - establishing a set of criteria, and then evaluating.

The original instrument was intended for all grade levels, but since a performance which is routine and low level when well learned may be high level while being learned—reading, for example—a new version of the instrument was developed for kindergarten or first grade.¹⁹

Although it is sometimes believed that the higher levels are more appropriate for older pupils than for younger, experience indicates that this is not true. The difference, of course, is that simpler materials and concepts are involved at lower grade levels. The development of a Piagetian concept such as conservation would fall at the level of synthesis, and the discussion that accompanies a story or a reading lesson may deal with questions such as "What else might Jimmy have done?" (synthesis), or "Would it have been better if Jimmy had done something different? Why?" (evaluation).

Reciprocal Category System (RCS). This is an extension of the Flanders System of Interaction Analysis, which is one of the most widely used instruments and whose well-known 10 × 10 matrix first permitted analysis of sequences.²⁰

In the modification by Ober, Wood, and Roberts, the seven teacher categories were expanded to nine.²¹ Teacher lecture is divided into that which is responsive to pupils, and that which is teacher initiated; the category of teacher criticism is divided into a category for correction without criticism, and one with criticism. Category 10 remains silence and confusion as before. Each of the categories is redefined, however, so that it can be used for pupil talk as well as for teacher talk. That is, teacher amplification of a pupil's idea is coded as a 3; a pupil amplification is a 13. Each category is changed from a teacher category to a pupil category by adding a "1" as the first digit. The observer, then, learns nine categories as he did with the Flanders System but has 18 to work with and, as a consequence, the same variety of pupil talk is recorded as teacher talk. Practically speaking, this modification offers greater richness of data at little increase in the complexity of the observer's task. A strength of the matrix procedure initiated in Flanders' work is the capturing, one step at a time, of the sequence of occurrences by the way the categories are tabulated. An interesting feature of the RCS is that it can be broken down into four submatrices: teacher-teacher talk, teacher-pupil talk, pupil-teacher talk, and pupil-pupil talk.

Observation Schedule and Record, Form 5, Verbal (OSCAR5v). This is the current revision of an instrument which has a history beginning in 1958.²² Without question, the various forms of OSCAR and the Flanders System of Interaction Analysis have been the two instruments used more than any

others by researchers.

The current revision of OScAR differs sharply from the original, and codes interactions as sequences. It uses 18 categories—four for pupil utterances and 14 for teacher utterances. Since six of the teacher categories are dual-purpose, however, there are, in effect, 20 for teacher behavior. The instrument provides for recording interactions by pairs of items which record either pupil utterance and teacher evaluation or response, or teacher question and the teacher's response to the pupil's response.

The fact that both entry to and exit from the interchange may vary means that the number of combinations—the number of specific interchanges—is large, and although the primary focus of the instrument is substantive, social-emotional behavior in management is also recorded.

One aspect of the instrument which seems likely to be important in a variety of settings and with a variety of objectives in PBTE is the ingenious system for recording interchanges—which not only records a teacher behavior, but also its antecedent. It seems likely that PBTE could use such a procedure for coding sequence with a variety of kinds of content.

Personal Record of School Experience (PROSE). This instrument is a system for recording the behavior and experience of individual young children.²³ The method used is time sampling. For a "cycle" of five observations, on signal from a timer, every 30 seconds the observer records the activity of the child at that moment. A series of 11 "words" are used to describe the activity; words 1-3 describe the adult-child contacts, with 1 recording the activity of the adult; words 4 and 5 record child-child interactions, with 4 the target child and 5 the child with whom he is interacting; word 6 records the level of the activity of the target child in the previous words, in terms of physical contact, materials, or whether verbal or non-verbal communication occurred; word 7 contrasts the sex and ethnic group of the target child with the other person; words 8, 9, 10, and 11 describe the appropriateness of the activity, the nature of the physical activity (movement), the nature of other activity (work, fantasy, etc.), and whether the activity was structured by a teacher or by a child.

After five observations (a "cycle") have been coded, the observer records the child's context in terms of the general activities of the classroom as a whole, other adults, and the instructional content.

Classroom Observation Instrument (COI). This instrument is the one which Stanford Research Institute has used to represent the major programs of planned variation in Headstart and Follow Through.²⁴ The instrument provides for a physical description of the classroom (recorded at the end of the day) in which numbers of adults and children are recorded, along with physical aspects such as noise level, lighting, furniture arrangement, and whether pupils select work groups. The observation begins with a Classroom Check List, which records the grouping patterns, materials in use, and adult responsibility. The Five-Minute Observation records four items of information for each of 76 frames: Who, To Whom, How, and What. The first two identify the initiator and receiver; What includes such items as command or request, open-end question, response, or instruction. How includes such items as nonverbal, motion, happy, or positive. The information recorded in each frame is summarized as a *sentence* which describes an event, and

provision is made for simultaneous or for recurring events.

Since there are thousands of possible combinations for a sentence, it is necessary to combine them into variables for further analysis.

Other Systems. There are hundreds of systems which have been developed, and many are reviewed by Rosenshine and Furst.²⁵ The most extensive collection of instruments, with annotations, is the series of volumes by Simon and Boyer.²⁶

A Final Comment

When one reviews the major points of this paper, the problems in classroom observation seem overwhelming. We have often approached teaching-learning as a simple, linear relationship while it is neither simple nor linear. The statistical problems are great; the problem of measuring outcomes for pupils is complex; and the measurement of classroom behavior is even more difficult.

Yet there is no alternative to facing these problems, because the most important part of schools is what is happening in the classrooms.

FOOTNOTES

¹ B. Rosenshine, *Teaching Behaviors and Student Achievement*, (London, England: International Association for the Evaluation of Educational Achievement, 1971)

² R. W. Heath and M. A. Nielson, "The Myth of Performance Based Teacher Education," paper presented at the American Educational Research Association (New Orleans, Louisiana: February, 1973), 16

³ J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, (New York: Academic Press, 1969)

⁴ R. S. Soar, "Optimum Teacher-Pupil Interaction for Pupil Growth," *Educational Leadership Research Supplement 1*, (1968), 275-280.

⁵ J. Gordon, ed., *Early Childhood Education*, (Chicago, Illinois: National Society for the Study of Education, 1972) "An Empirical Analysis of Selected Follow Through Programs. An Example of a Process Approach to Evaluation," by R. S. and R. M. Soar, 229-259.

⁶ R. S. Soar and R. M. Soar, *Classroom Behavior, Pupil Characteristics, and Pupil Growth for the School Year and for the Summer*, (Gainesville, Florida: University of Florida, grant no 5R01MH15891, and 5R01MH15626, Washington, D.C.: National Institute for Mental Health, Department HEW, 1973)

⁷ Ibid

⁸ R. S. Soar, "Optimum Teacher-Pupil Interaction for Pupil Growth," 275-280.

⁹ R. S. and R. M. Soar, "An Empirical Analysis of Selected Follow Through Programs," 1972, 229-259

¹⁰ R. S. Soar and R. M. Soar, *Classroom Behavior, Pupil Characteristics, and Pupil Growth for the School Year and for the Summer*, 1973

¹¹ J. R. Squire, ed., *A New Look at Progressive Education*, (Washington, D.C., Association for Supervision and Curriculum Development, 1972) R. S. Soar, "Teacher Pupil Interaction," Chapter V, 166-204

¹² R. S. Soar, *An Integrative Approach to Classroom Learning*, NIMH Project Numbers 5-R11MH01096 to the University of South Carolina, and 7-R11MH02045 to Temple University, (Philadelphia, Pennsylvania, 1966.)

¹³ H. Taba, S. Levine, and F. F. Elzey, *Thinking in Elementary School Children*, Coop Res. Project No. 1574, O E U.S. Department HEW (San Francisco, California: San Francisco State College, 1964.)

¹⁴ N. A. Flanders, "The Changing Base of Performance-Based Teaching," (*Phi Delta Kappa*, vol. 55, 1974), 312-315.

- ¹¹ R S Soar, R M Soar, and M Ragosta. *Florida Climate and Control System. Observers Manual.* (Gainesville, Florida University of Florida, 1971)
- ¹² R S Soar and R M Soar. *An Empirical Analysis of Selected Follow Through Programs: An Example of a Process Approach to Evaluation.* 229-259.
- ¹³ R S Soar. *An Integrative Approach to Classroom Learning.* 1966
- ¹⁴ D Fowler. "Relation of Teacher Personality Characteristics and Attitudes to Teacher-Pupil Rapport and Emotional Climate in the Elementary Classroom." (Ph D dissertation, University of South Carolina, 1966)
- ¹⁵ B B Brown, *The Experimental Mind in Education.* (New York Harper and Row, 1968).
- ¹⁶ B S Bloom, et al., ed., *Taxonomy of Educational Objectives Handbook I.* (New York McKay, 1956)
- ¹⁷ N M Sanders, *Classroom Questions. What Kind?* (New York Harper and Row, 1966)
- ¹⁸ B B Brown, R Lober, R S Soar, and J. N. Webb, *The Florida Taxonomy of Cognitive Behavior.* (Gainesville, Florida Institute for Development of Human Resources, University of Florida, 1972)
- ¹⁹ R S Soar, M Ragosta and R M Soar, *Florida Taxonomy of Cognitive Behavior (K-1 version)* (Gainesville, Florida University of Florida, undated)
- ²⁰ N A Flanders, *Teacher Influence, Pupil Attitudes, and Achievement.* Coop Res Monograph No 12. OE 25040. (Washington, D C U S Department of HEW, 1965) N.A Flanders, *Analyzing Teacher Behavior.* (Reading, Mass. Addison-Wesley, 1970)
- N A Flanders, *Teacher Influence, Pupil Attitudes, and Achievement.* Coop Res. Monograph No 12. OE-25040. (Washington, C C U.S Department of HEW, 1965)
- ²¹ R L Ober, S E Wood, and A Roberts, *An Instructional Manual for the Reciprocal Category System* Based on a paper read at the annual meeting of the American Educational Research Association, (Chicago, Illinois, February, 8-10, 1968.)
- ²² D M Medley, *The Observation, Schedule and Record. AV Coders Manual.* (Charlottesville, Virginia University of Virginia, 1974)
- ²³ D M Medley, T J Quick, C G. Schuck, N P Ames, M F Carrico, and D. G. Massey, *The Personal Record of School Experience.* (Charlottesville, Virginia University of Virginia, 1973)
- ²⁴ J A Stallings, *Follow Through Program Classroom Observation Evaluation, 1971-1972.* Stanford Research Institute, Contract #OEC-0-8522480-4366(100). (Washington, D C Bureau of Elementary and Secondary Education, Office of Education, Department of HEW)
- ²⁵ R M W Travers, ed, *Second Handbook of Research on Teaching.* (Chicago, Illinois. Rand McNally, 1973) "The Use of Direct Observation to Study Teaching," by B. Rosenshine and N Furst, 122-183
- ²⁶ A Simon and E G. Boyer, eds., *Mirrors of Behavior: An Anthology of Classroom Observer Instruments.* (Philadelphia, Pennsylvania. Research for Better Schools, Vols. 1-6, 1967, Vols 7-14, and Summary 1970)

8 Inservice • Needs Assessment • Competency Based Teacher Education

Gilbert F. Shearron

Public schooling probably is the only large-scale enterprise in this country that does not provide for systematic updating of the skills and abilities of its employees and for payment of the costs involved. Teachers, we presume, can acquire markedly different ones through some process of osmosis.¹

Most persons and organizations who give serious thought to inservice education (staff development seems to be the current term) would probably agree with the above quote from John Goodlad. One of the difficulties in inservice education has been that it only provides information to teachers—information about what someone else (usually a consultant or college professor or textbook) thought or information shared between teachers. Little or no consideration has been given to assisting teachers in developing and practicing new teaching skills in the classroom. And little attention has been given to assessing the effectiveness of what the teacher does, and what the pupils learn.

Bob Bhaerman has summarized inservice education by saying that often it has been long on service and short on education. It has been too fragmented, without integrated activities based upon assessed priority needs, insufficiently supported by budget, and too insignificant to leave a marked and continuing impact upon teachers and programs.

Inservice education is a term that has various interpretations. For purposes of this paper, it is defined as the continuous training that a teacher needs as a result of being assigned to certain teaching situations. School districts have unique needs. Sometimes schools within school districts and classrooms within these schools also have unique needs. Few, if any,

preservice teacher education programs make the claim that their graduates have all of the teaching skills necessary to function in any circumstance. Therefore, inservice education should focus first on developing those additional teaching skills needed to implement the regular program prescribed by the school district. Secondly, the inservice program should focus on providing knowledge and teaching skills that result from change in both curricula and pupil population.

For example, in the early 1960's, teachers who were asked to teach "new math" might have needed knowledge and new methodological skills. As another example, if a school district adapts a diagnostic, prescriptive teaching strategy, then the teacher may need knowledge of diagnostic techniques as well as practice in the application of diagnostic procedures. Thirdly, inservice education aims at the need for the continuous improvement of educational personnel.

Inservice education cannot be effective without considering what needs teachers have. The position advanced in this paper is that a viable inservice education program begins with a needs assessment procedure that requires the following:

1. The identification of teaching skills needed in a particular teaching situation.
2. The development of a procedure for determining teacher needs in relation to the identified teaching skills.
3. A procedure for assisting teachers to acquire and improve his or her teaching skills.

This paper speaks to these three points. It discusses and outlines a strategy based on Competency Based Teacher Education (CBTE).

This paper does *not* address itself to the knowledge teachers might need. It focuses on improving teaching skills. The paper also does *not* deal with the question of who is responsible for teacher growth. While this is a most crucial issue, it is not within the scope of this discourse. The paper is limited to a needs assessment strategy *only* for classroom teachers.

Why CBTE?

Competency Based Teacher Education is a data-based system for training teachers. CBTE precisely specifies that teachers must exhibit those competencies assumed to promote pupil learning, and/or demonstrate their ability to promote desirable pupil learning.³ Instruction in CBTE focuses directly on the development of those competencies thought to promote pupil learning. It is the opinion of the author that CBTE is a viable approach to needs assessment because it is a data-based system. William Drummond points out that in a CBTE program, data are collected and used to make adjustments and changes in what and how things are done. Data are not collected for punitive reasons.⁴ It is crucial that needs assessment be thought of as a procedure for improvement, and not to "get teachers."

The basic elements in CBTE provide a way to develop a needs assessment system that addresses itself to the three requirements for such a system mentioned earlier in this paper. Stanley Elam identifies the essential elements of CBTE as:

1. Teaching competencies to be demonstrated are role derived, specified in behavioral terms, and made public.
2. Assessment criteria are competency based, specify mastery levels, and are made public.
3. Assessment requires performance as prime evidence, takes knowledge into account.
4. Student's progress rate depends upon demonstrated competency.
5. Instructional program facilitates development and evaluation of specific competencies.⁵

The Elam essential characteristics are perhaps more often thought of in the preservice realm. They can, however, be applied to the needs determination-inservice area. The position of this paper is that *conceptually*, CBTE is more easily adapted to inservice education than to preservice education.

Elam's five essentials fit nicely into a needs assessment scheme. Items one and two offer the opportunity to identify teaching skills in particular situations. Items three and four provide direction for establishing a procedure to determine needs, and the final item addresses itself to the improvement of teaching skills.

The systematic data-based approach utilized in CBTE provides a means to tie needs assessment and inservice together with other parts of the school program. Robert Houston and Robert Howsam point out that CBTE is planning in systemic terms, dealing simultaneously with the elements that comprise a total system.⁶ A systematic data-based approach can make possible the continuous updating of teaching skills. The use of a systems approach, however, requires that specific decisions be made as to what is to be accomplished. It also requires that problems in inservice teacher education can neither be resolved nor exist in isolation from the total program of the school district. As an example of this, one might consider that if the school district is unwilling to provide released time or extra compensation for a teacher to improve a skill that the district itself says he or she must have, then there is little likelihood that the new skill will be developed.

The Politics of Assessment

The assessment of the needs of school personnel can be threatening to those who are being assessed. The threat may or may not be implied. A decision to identify teaching skills presumed to be necessary for existing and new programs can become a matter of great concern to some individuals charged with carrying out the program. Organizations representing the teachers become concerned that the assessment will be used as a device to dismiss and retain teachers, rather than as a device to improve teaching.

In addition to concern about assessment, the introduction of CBTE at the inservice level can become a highly partisan issue. Unfortunately, CBTE often elicits an emotional response rather than reasoned consideration. Theodore Andrews believes that teachers will demand a significant role in any competency based system that includes inservice education.⁷ A few of the political questions to be dealt with are: Who will determine what teaching skills are required? Who will determine which teachers need which skills? Who will determine what learning activities are appropriate to

develop certain skills?

The position put forth in this paper is that the design of a needs assessment system in a CBTE framework must consider, along with the conceptual strategy, a corresponding political strategy. The integration of these two strategies calls for a systematic approach to planning and implementation.

A Proposed Strategy

The strategy offered in this paper is an attempt to combine the conceptual aspects of needs assessment with political considerations. Generally, the conceptual notions come from attempts to apply the essential elements of CBTE to needs assessment. The political implications come from pilot efforts to use these conceptual notions in needs assessment and inservice programs. The pilot work has been done in three school districts: a large urban district, a middle-class suburban district, and a small rural district.

Generally, the strategy begins with the determination of some of the teaching skills needed in a particular school district. Once needs have been established, assessment procedures are developed to determine where teaching personnel are in relation to these needs. Finally, the needs assessment strategy leads to inservice education opportunities.

There are two assumptions that provide the foundation for this strategy:

1. Needs assessment is based on the improvement of teaching skills.
2. Those who are to be assessed should be active in the development of the assessment procedure.

Identification of Needs

The identification of needed teaching skills is the first step in the strategy. By utilizing a CBTE approach, teaching skills are identified as competencies that teachers should have in order to carry out the program of the school district. The school district has to make a decision as to how the teaching function should be analyzed (how competencies should be determined). Later, we will describe two approaches used by two different school districts to determine this.

The next question to be considered is: who is to analyze the teaching function? Most teachers in the field are inclined to believe that they know more than anyone else about what a teacher does. Researchers and those who attempt to project what a teacher should be usually feel that they know more about the role of a teacher. However, inservice teachers are not likely to be willing to have someone else analyze the teaching function, and then have the administration hold them responsible for these functions. We have found that teachers are willing to listen to the advice of so-called experts, but they intend to reserve the final decisions for themselves.

An illustration of a role the "experts" might play would be in helping the teachers to clarify ideas. Suppose that one of the competencies needed is the skill of questioning. Here the teachers might be aided by looking at a number of alternative questioning strategies which would help to define

what types of questioning skills are appropriate, how they might be assessed, and what types of training activities are available.

One approach to developing inservice competencies has been tried by the Atlanta Public Schools. A comprehensive revision of their elementary curriculum was carried out in the early 1970's. Five areas of literacy were developed for all pupils: personal, career, aesthetic, social, and intellectual. Major ideas, concepts, and objectives were identified for each of these areas. In addition to the five areas, a basic instructional strategy was developed to carry out the curriculum.⁸ The curriculum was developed by teachers, community, professional organizations, etc.; then the curriculum was piloted in 10 elementary schools. After one year of piloting, one of the pilot schools was selected to develop competencies thought to be necessary to implement this curriculum.

The teaching staff of the school, university personnel, plus curriculum and supervisory personnel from the district office began to systematically determine what competencies were needed to implement the curriculum. As an illustration, Figure 1 depicts the instructional strategy to be employed in carrying out the curriculum. The teaching staff analyzed the strategy, applied their experience in piloting the curriculum for one year, and began to determine what skills a teacher needed. The strategy (see Figure 8.1), as you see, begins with "Find what student knows." This implied diagnostic skills. The five areas of literacy referred to earlier indicate that diagnostic skills need to go beyond the intellectual into the social, aesthetic, and personal areas. There are, of course, additional competencies needed to utilize the prescribed strategy.

The Atlanta approach might be classified as a curricula approach to analyzing the teaching function. They took the basic curriculum with its instructional strategy and from this, inferred competencies for teaching. The strategy employed in Atlanta was involvement of the teachers (among other groups) in developing the curriculum. Teachers then had the opportunity to define the teaching skills needed for implementation. While there is still, perhaps, some apprehension on the part of some teachers, this should be reduced because of the involvement of the teachers from the beginning.

Another effort to develop competencies for purposes of assessment was done by the Mentor, Ohio Exempted Village School District. The Mentor project utilized differentiated staffing within two elementary schools using open-space settings. Within these two schools, professional and paraprofessional staff members were employed in a role hierarchy (i.e., master teacher-instructional strategist, staff teacher, assistant to the teacher, and intern). The project focused attention on the need to enhance instructional opportunities for pupils by developing a systematic approach for continuous staff growth based on individual as well as differentiated staff needs. Specifically, the project focused on specifying competencies for teachers in a differentiated staffing pattern. These competencies were to be used in an assessment procedure to determine what types of staff development activities were necessary to enable the teaching staff to develop and maintain the desired level of proficiency.¹⁰

The strategy utilized in Mentor was initially to have the personnel involved

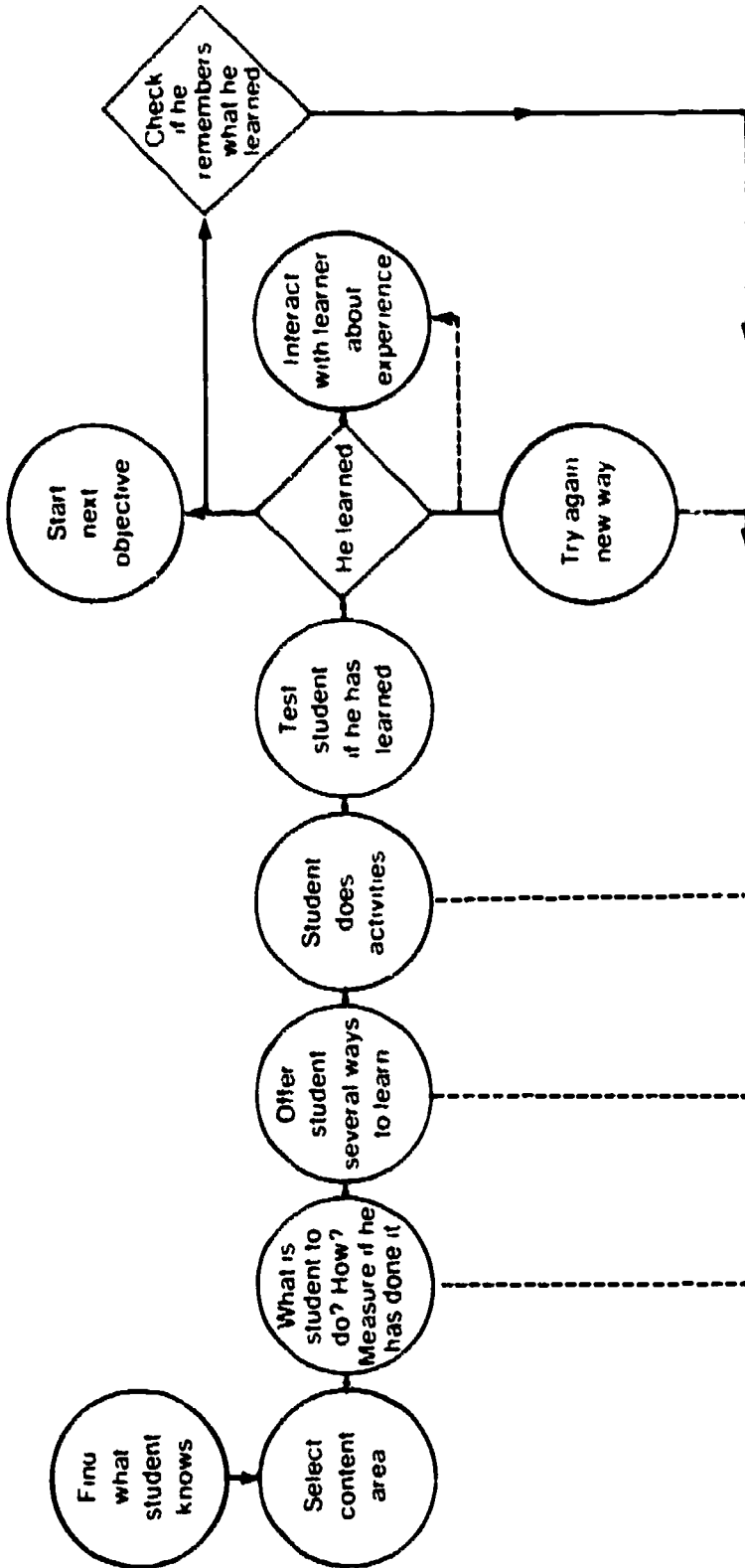


Figure 8.1
INSTRUCTIONAL STRATEGY UTILIZED IN
ATLANTA'S ELEMENTARY CURRICULUM REVISION

react to an instrument developed to measure perceptions of competency levels within each school. The statements utilized in the instrument were drawn from the literature on differentiated staffing and from current operational job descriptions. Staff members were asked to react to the degree to which the statements were operational in their buildings. They were also asked to indicate the importance of each statement. One of the statements on the instrument was "Teachers in my building show evidence of successful planning of learning activities." Statements such as this provided a basis of discussion which led to the specification of competencies for a differentiated staffing pattern.¹¹

The Mentor approach can perhaps best be described as a speculative approach to analyzing the teaching function. The instructional staff, with some structure present in the instrument, speculated on what teachers do within a differentiated staffing pattern. The political strategy is present because those who are to be affected were making the decisions.

There are, of course, other approaches to developing competencies for inservice teachers. The two illustrations were used because of the attempt to put together conceptual and political strategies. In both cases, once the competencies were identified they provided the baseline data that describes where the school district wants to go and what a teacher should be able to do.

It is interesting to note that in both the Atlanta and Mentor efforts, neither group identified as a teaching competence any of the strongest variables in the relationship between teacher behavior and student learning identified by Barak Rosenshine and Norma Furst.¹² This would indicate the lack of inputs by researchers into determining competencies for teachers in the two districts. Consequently, we would suggest that part of any identification of teaching skills be preceded by opportunities for consideration of research on teaching. Most teachers have not had the opportunity to consider the theoretical and research aspects of teaching. Although the decision for determining the specific competencies should be the responsibility of teachers, they need to be exposed to the study of teaching.

Assessing the Needs

The assessment of needs requires that some type of scheme be developed that can describe where a teacher is in terms of the school district's requirements and can also provide an opportunity to chart a direction for improvement. In developing assessment procedures, our experience has been that you begin with the assumption that persons who have a certificate and have been employed by a school district are competent in the opinion of those who granted the certification and those who employ. This assumption may be odious to some readers; but nevertheless, it is difficult to start with the assumption that some teachers are competent and others incompetent because competence without some data base is a relative thing. A teacher with a certificate and a job has as much claim to competence as anyone else. It also is, in our opinion, unwise to start with the proposition that assessing needs is a basis for finding out who is and who is not competent. It has been our experience in the three school districts

that when we start from a base that everyone is presumed competent, there is less reluctance on the part of teachers to participate.

W. Robert Houston uses the illustration of the violin when discussing competence. He points out that the violin soloist at the symphony must be able to read music, properly handle the bow, tune the instrument, and have a stage presence. But so must the beginner in the seventh grade concert. The differences between the violin soloist and the seventh grader lie in the criteria which are acceptable for an adequate performance.¹³

Houston then points out the parallel in teaching. The beginning teacher may perform adequately in asking higher order questions, establishing sets, and writing criterion-referenced objectives, but might not be able to integrate those skills and employ them in given circumstances. The beginning teacher may be adequate the first year on the job, but after a few years of experience, she should be able to integrate the skills and employ them appropriately.¹⁴ In the Mentor model referred to earlier in this paper, the Master Teacher instructional strategist certainly should be able to more effectively utilize teaching strategies than the staff teacher.

The differences in levels of competence and the reluctance by teachers to be associated with incompetence led us to utilize a continuum of proficiency for each competency, the definition of proficiency being to advance or move forward. By utilizing the continuum, we have the opportunity to allow movement through inservice education toward new levels of proficiency. Theoretically, one can always become more proficient. The idea of continuous improvement of one's skills is difficult for any organized profession to be against. Point one on the continuum can be the position where one enters the profession or where minimum proficiencies are exhibited. The last point of the continuum can be the ultimate in teaching or the place where a teacher uses special skills in particular situations.

At each point on the continuum, a set of criteria must be developed that is acceptable at that level. A special set of criteria is needed for the violin soloist and a special set for the seventh grader. Figure 8.2 illustrates how a continuum of proficiency might look.

The criteria or indicators of proficiency help us to operationalize the competency. They provide the precise criteria and conditions under which the competency will be judged. In a needs assessment system, the criteria not only help in the identification of where the teacher is, but they will aid in setting directions for improvement. For every competency, representative indicators of proficiency should be developed. It is the position of this paper that they should be developed along a proficiency continuum. We use representative indicators because, in our opinion, the same indicators are not necessarily acceptable in all situations.

Teachers also feel that because of variation in conditions, the same indicators are not appropriate for every situation. For example, learning outcomes being related to pupil entry behaviors may not always be possible or desirable. Available data on pupil learning behavior may be limited only to the cognitive. Or the lesson being taught may focus on experiences that someone thinks are desirable for pupils to have. If the purpose is to experience something, then pupil entry behaviors may not be important.

Chances are that the indicators of proficiency may be appropriate for

Area of Competency: Planning	Statement of Competence			Teacher organizes lesson to facilitate learning
1	2	3	4	5
Proficiency Expected for Inexperienced Teachers	Proficiency Expected for Experienced Teachers			Proficiency That Extends the Effectiveness
Representative Indicators				
1. Learning outcomes stated behaviorally.	1. Learning outcomes are related to pupil entry behaviors.	1. Learning outcomes based on psychological and sociological principles.		
2. Conditions of each outcome stated in clear behavioral terminology.	2. Plan is flexible and dependent upon pupil feedback.	2. Individual learning outcomes for each pupil are stated.		
3. Achievement variables stated in learning outcomes.	3. Learning outcomes can be assessed.	3. Information feedback loops are included in the plan.		
4. Affective variables included in learning outcomes.	4. Materials appropriate to each pupil.	4. Materials are selected that are appropriate to the learning outcomes.		
5. Instructional materials clearly identified.	5. Assessment checkpoints are noted throughout the plan.			
6. Summative evaluation included.				

AN ILLUSTRATION OF A NEEDS ASSESSMENT PROFICIENCY CONTINUUM

Figure 8.2

some teachers in some classes, but they are certainly not appropriate for every teacher in every situation. Therefore, it is our feeling that those involved in the needs assessment should have the opportunity to negotiate additional indicators of proficiency either for purposes of assessment or for determining appropriate directions for his personal inservice education.

Negotiation becomes a critical feature of the proposed strategy. Whom the teacher negotiates with will be discussed in the succeeding paragraphs. But the politics of negotiations gives the teacher the opportunity to have his views plus his unique teaching situation considered. Conceptually, there are, at this point in time, few people who are willing to state categorically that they can specify all of the conditions and criteria under which one can be stamped *competent*.

The negotiation of proficiency also allows for dialogue among the teacher and those who participate in the assessment process. The beginning teacher can be assisted in considering what are the most appropriate skills to be developed immediately while the experienced teacher can be helped to establish some sort of priority. Moreover, it is also useful to have someone else's opinion in determining where one ranks on a proficiency continuum. Finally, negotiation makes the individual being assessed active in the entire needs assessment effort.

Who Assesses?

Negotiation implies that self-assessment be part of any needs assessment procedure. Our experience has been that most teachers express a desire to use self-assessment for one of the following reasons: (1) They hypothesize that they are not likely to change their performance unless they see a discrepancy between what they want to achieve and what they are actually achieving. (2) They feel that they know better than anyone else about their particular teaching situation.

John McNeil and James Popham point out that there is a tendency when using self-assessment for teachers to overrate themselves. Further, there are negligible relationships of self-assessment with other criteria such as student ratings and measures of student gain. McNeil and Popham also state that most teachers seem to criticize the superficial aspects of their teaching, personal mannerisms, appearance, etc., rather than the relevant aspects of their work.¹⁵

Still the political expedient remains. We believe that at this point in time, self-assessment will probably have to be part of any assessment procedure that carries teacher approval. However, to be effective in self-assessment, teachers should be trained to focus on the relevant aspects of their work, especially in judging the consequences of their behavior on pupils. Furthermore, teachers must have opportunities to practice these skills. There is competence in the application of self-assessment skills just as there is in asking questions. If self-assessment skills plus proper attitudes toward needs assessment can be built, then the acceptance of a needs assessment system can be enhanced. It is our opinion that self-assessment and the political strategy are complimentary.

A needs assessment system also should be related to the administrative hierarchy of the school district. There should be some type of subordinate-superordinate relationship. If the teacher is in the subordinate role, then the role of superordinate might be played by a principal, supervisor, or in a differentiated staffing pattern by a team leader. The superordinate, while perhaps posing a threat in some instances, does have some responsibility for seeing that programs are carried out. If needs assessment is to be based on improvement, then the superordinate has responsibility for this, being in a position to make the inservice effort more responsive to the teachers' needs. It is the position of this paper that through the use of indicators of proficiency, with the teacher and superordinate negotiating the indicators, there is less possibility of threat to the teacher. The possibility of dialogue for improvement and understanding is also expanded.

As an illustration, suppose that a competency for a teacher is in the area of diagnosis. Teacher A has no particular skills in diagnosing pupils. Teacher A and the principal agree that the teacher should focus first on diagnosing the intellectual abilities of pupils. They agree that one of the criteria to be used in assessing intellectual abilities is the use of teacher-made diagnostic tests for groups and individuals. In order to become more proficient, Teacher A is to develop a skill in building tests that can identify where pupils are on specific skill continuums in reading and mathematics. Inservice activities should then be recommended by the principal that will help Teacher A to learn how to construct and administer diagnostic tests in reading and mathematics. Once the teacher develops the skill and utilizes it in the classroom, then there would be opportunity to develop further diagnostic skills that would continue to make the teacher more proficient.

We believe that teachers and administrators can work together in identifying needs if it is for purposes of improvement. Obviously, there are the human requirements of trust and consideration that must be present in the relationship for this or any other working relationship to succeed. We are hypothesizing that if a subordinate and a superordinate agree on an area of improvement, then precisely define what the improvement will be, there is a good possibility that improvement can be assessed by both parties. Again it is also important to consider that the administrative hierarchy be responsive to the needs of the teacher. A hierarchy is likely to react more readily if it is represented in the process.

Summary

This paper has presented a proposed strategy for developing and utilizing a needs assessment strategy based on Competency Based Teacher Education. The strategy takes into account both conceptual and political considerations. It comes from attempts to combine some of the conceptual elements of CBTE with pilot efforts to use these conceptual notions in school districts. Some of the suggested procedures are as follows:

1. Teachers should make the final decisions on needed competencies to perform effectively in a school district. However, teachers should

- have opportunities to consider the theoretical aspects of teaching, including research findings prior to making decisions about needed competencies.
2. The assessment procedure should begin with the assumption that needs assessment is for the improvement of teaching. Therefore, we begin with the notion that teachers are competent, and that the purpose of the needs assessment is to help them become more proficient.
 3. A continuum of proficiency for each competency should be developed in order to allow for continuous teacher growth. The indicators of proficiency on this continuum should be negotiated by the assessor and the teacher.
 4. Self assessment should be part of the needs assessment procedure. However, self-assessment skills need to be developed in order to make this a successful venture.
 5. The administrative hierarchy of the school district should be involved in the assessment procedure.

FOOTNOTES

- ¹ John I. Goodlad, "The Schools vs. Education," *Saturday Review*, (April 19, 1969), 59-61
- ² Bob Bhaerman, *A Paradigm for Accountability*, American Federation of Teachers Quest Paper, (No. 12, August, 1970), 3
- ³ Gilbert F. Shearron and Charles E. Johnson, "A CBTE Program in Action: University of Georgia," *Journal of Teacher Education*, 24 (Fall, 1973), 187
- ⁴ William H. Drummond, "Comments on Achieving the Potential of PBTE," Appendix B-3, *Achieving the Potential of Performance-Based Teacher Education Recommendations*, (Washington D.C. American Association of Colleges for Teacher Education, December, 1971), 39
- ⁵ Stanley Elam, *Performance-Based Teacher Education: What is the State of the Art?* (Washington, D.C. American Association of Colleges for Teacher Education, December, 1971), 18
- ⁶ W. Robert Houston and Robert B. Howsam, *Competency-Based Teacher Education: Progress, Problems, and Prospects* (Chicago: Science Research Associates, Inc., 1972), 11
- ⁷ Theodore E. Andrews, *Atlanta or Atlantis?* A publication of the Multi-State Consortium on Performance Based Teacher Education
- ⁸ Atlanta's Elementary Curriculum Revision Project, "Your Schools, a publication at the Atlanta Public Schools, (April, 1973), 3
- ⁹ *Ibid*
- ¹⁰ *An Assessment System Utilizing Teacher Competencies for Differentiated Staffing*, Developed by Mentor Ohio Exempted Village School District and The Center for Competency Based Education, University of Georgia, (Athens, Georgia, February 1974)
- ¹¹ *Model Teacher Education and Differentiated Staff Assessment*, Mentor Ohio Exempted Village School District, ESEA Title III Project (45-71-208-2)
- ¹² Barak Rosenshine and Norma Furst, "Research on Teacher Performance Criteria," Chapter 3, in B. O. Smith (ed.) *Research in Teacher Education*, (Englewood Cliffs, N. J. Prentice Hall, Inc., 1971)
- ¹³ W. Robert Houston, *Strategies and Resources for Developing a Competency Based Teacher Education Program* (New York: State Education Department and Multi-State Consortium on Performance-Based Teacher Education), 21
- ¹⁴ *Ibid*, 21-22
- ¹⁵ John D. McNeil and W. James Popham, "The Assessment of Teacher Competence," *Second Handbook of Research on Teaching* (Chicago: Rand McNally College Publishing Company, 1973), 231-2

9 Assessing Teacher Competence

**Beatrice A. Ward
Greta Morine
David C. Berliner**

Assessment of teacher performance may be conducted for a variety of reasons, such as selection for entrance into preservice training, identification of the types of training most needed by a preservice and/or inservice teacher, awarding of a certificate to teach, and research into effects of teacher use of particular skills.

For the purposes of this paper, we are focusing upon assessment as a means for selecting training options. We have made this choice for two reasons. First, the body of knowledge in teacher effectiveness research at the present time is too limited and too subject to question to justify using assessment data for selecting preservice candidates or granting licenses to teach.¹ We do not as yet know enough about which aspects of teaching are essential for affecting student performance positively to use specific levels of teacher performance as "go" or "no go" criteria for entry into or promotion within the profession.

Yet much is known about measurement of teacher performance. Large numbers of evaluative studies report data describing typical levels of teacher performance before and after training.² To ignore these data while waiting for completion of effectiveness research is, in our opinion, as great a mistake as using them for hiring or firing purposes. Thus, our second reason for concentrating upon assessment for training purposes is to illustrate how the current state-of-the-art in the study of teaching can be productively applied to the design of a self-improvement program for an individual teacher or a group of teachers.

Within the spectrum of performance based teacher education, several training approaches have been developed. The assessment techniques described herein illustrate two diverse points on the continuum of ap-

proaches. The skill diagnosis approach represents a behavioristic view of assessment, the inquiry approach, a humanistic view. Three critical differences in the approaches are (a) the data base from which decisions are made regarding the need for training, (b) the individuals responsible for designing the assessment event, and (c) the type of evaluative feedback provided.

Both approaches are under development at the Far West Laboratory for Educational Research and Development. Work on the performance tests of teaching is in cooperation with National Teacher Corps and Corps sites at Kansas State Teachers College, Emporia, Kansas; Buffalo State University College, Buffalo, New York; and the Southern Consortium of Colleges and Universities, Mobile, Alabama. The inquiry approach is being developed in cooperation with the Oakland Unified School District, Oakland, California, under the auspices of the Bay Area Teacher Training Complex, sponsored by a grant from the U. S. Office of Education.

A Skill Diagnosis Approach

The skill diagnosis approach to analysis of competency in teaching competency is based on a belief that teacher trainers and researchers need to introduce a medical model into the way they examine teachers and their teaching practices. Within the medical profession, the model requires the monitoring of a number of potentially important systemic indicators, such as heart rate, patellar reflex, and blood pressure. Then clinical knowledge about the client is developed—his or her job requirements, temperament, previous history, etc. On the basis of these data, if certain indicators are out of tolerance, individualized prescriptions are made (e.g., lose weight, quit smoking, take medication four times a day).

Teacher trainers and researchers also need to develop ways to monitor the systemic functioning of their clients, perhaps checking their warmth instead of their temperature and their response to pupils rather than to light. Then, with clinical insight derived from information about the teacher's experience and interests—and perhaps the age and social class of the students to be taught—individualized prescriptions could be prepared, such as: "Take Module 014 from the Houston Module Bank for learning reinforcement techniques to use with young inner-city children."

How to find the indicators for diagnosis and prescription of this kind is the concern of the skill diagnosis approach to assessment.

Knowing "What" and Knowing "How"

Multiple choice or essay tests are easy to develop and grade, can sometimes even be reliable, and cost very little per client tested. One can readily assess knowledge about *what* discovery-teaching is, *what* one should do with obscene notes, *what* the definition of contingent reinforcement is, etc. This knowledge of "what" may be important, but its relationship to knowing *how* to teach is quite indeterminate. A recent view correlating a highly reliable test of knowledge about teaching with various measures of teaching performance informs us that virtually no relationship exists between the two variables.³ Indicators for examining teaching should not come from

paper-and-pencil tests which measure *what* is known. The indicators must be based on *how* teachers perform.

Observation in Natural Environments

Some believe, along with the many developers of observation instruments, that observations of how teachers perform in their classrooms is the way to assess teaching competency. But inclass observations might be very much like the annual visit to the classroom by the principal—an exercise in futility. In most philosophic systems, the "good" teacher is one who varies his or her behavior depending upon the child, the learning objective, the time of year, and a host of other variables. All these are influences on teacher *behavior designed to make the behavior unstable in situ*. Thus observation of teachers in natural environments may be an inappropriate way to develop reliable and valid indicators of teaching performance.

Performance Tests

In performance tests of teaching, where a controlled environment exists, indicators for skill diagnosis may be found that are similar in type to those used by our medical colleagues. In these diagnostic tests, a teacher trainee receives a statement of two or three objectives, some written material on a topic, and sample test items. The teacher is told to prepare to teach five or six students for 20 minutes. The teaching performance is video- or audio-taped. Frequency counts or ratings on as many as 50 variables are completed.

In order to develop such tests, three important elements must be considered:

Choosing variables. Thus far, the literature on teacher effectiveness (such as it is) along with the value systems of the participating researchers and teacher trainers have directed the selection of variables. Variables also have been picked to match those included in existing teacher training materials since it is important to examine the pharmacy, i.e., the Mini-courses, texts, protocols, simulations, etc., available for use in responding to a skill diagnosis before writing a prescription. In the performance tests, we try not to be like the doctor who diagnoses terminal cancer and offers no cure. We assess first if there is a cure, i.e., a training module, and then proceed to diagnose by examining some indicator to see what, if any, training need exists in the curable area.

Evaluating performance. The teacher's performance is evaluated against: (a) norms which have been accumulated from other trainees, (b) the performance of master teachers attempting the same lesson, and (c) criteria we have set for performance in a particular area. In light of each or all of these sources of information about teaching performance, a diagnosis is made. Feedback is provided to the trainees on each diagnosis. For example:

Variable 1: Teacher gives specific praise or encouragement

Variable Description: The teacher praises or encourages students and clearly states why praise or encouragement is given. Example—"That is

a good answer because . . ."

Performance Evaluation

- 1.) This evaluation occurred 3 times in your teaching.
- 2.) Based on norms and our criteria, 5 to 12 times is desirable.
- 3.) The diagnosis to be made is:
 - _____ Performance is within limits
 - _____ Performance is too high
 - X Performance is too low

Prescriptions: If performance is too high, see Nos. 2, 4, or 31.
If performance is too low, see Nos. 30.1, 32.

This form of feedback is continued for as many variables as desired. In each case, the trainee receives descriptive information and definitions of the variable, his actual performance, estimates of desirable performance, and prescriptions for remediation.

Prescription. The numbers in the prescription list refer to training materials which exist in the "pharmacy." For example, No. 30.1 may be the code for a lesson in a Minicourse that focuses on contingent reinforcement; No. 32 may refer to a protocol videotape that demonstrates the concept of "because extensions" in communication, or the trainee may be referred to a chapter in a text on the effects of praise and blame on performance. If the trainee's performance was too high, that is, if performance exceeded the levels considered desirable, the trainee might be directed to materials which teach about habituation, since praise will become ineffective if too much is given, because people tend to disregard it or question its genuineness.

For any teacher variable of interest, if an out-of-tolerance situation is detected in a performance test, one or more prescriptions can be recommended. In this way these tests are as diagnostic as the physician's examination. An individually prescribed program of training may be developed which allows a teacher to concentrate on areas where training will be most useful and skip areas where performance competency has already been demonstrated. Preliminary study of these performance tests has shown them to be reliable. Stability of performance seems to hold over parallel lessons and for the same lesson taught at adjacent times. This means that the performance tests, when thoroughly developed, can be used not only to plan training but also as objective measures of the effectiveness of competency based teacher training programs.

An Inquiry Approach

All approaches to assessing teacher competence share at least one common purpose—the improvement of instruction. The two approaches discussed here share a common orientation to attainment of that goal—the view that *formative* evaluation of teacher effectiveness is the most viable route. The inquiry approach differs from the diagnostic/prescriptive approach in actively pursuing the following additional purposes:

1. To provide a supportive climate in order to maximize change in teacher behavior.
2. To individualize the explicitly stated criteria on which assessments of teacher competence are based.
3. To provide teachers with a strategy for continuous self-assessment and self-improvement.

The inquiry approach to be described here had its roots in an analysis of the process of inquiring into one's own teaching behavior proposed by Flanders.⁴ This analysis was directly related to assessment of teacher competence in a later paper on the California Stull Act,⁵ which proposed teacher inquiry projects as one method of assessment. The comparative procedure which is an essential element of this approach was tested and found to be effective in getting teachers to be more specific in identifying training needs and more interested in continued skill training.⁶

Procedure for Inquiry Projects

A teacher embarking upon an inquiry project for the purpose of self-assessment would engage in the following steps:

1. Identify one or more instructional objectives and plan two alternative lessons to achieve the objective(s).

Complexity of plans may vary here. An instructional objective may be simple and immediately attainable, such as "to identify correctly six words as nouns, given a list of ten nouns and verbs." The objective may be more complex and long-term, such as, "to increase pupil skill in listening to and reacting to each other's comments." The alternative lessons for achieving the instructional objective may involve simple variation in use of one particular teaching skill, such as praising pupils for every correct answer in one lesson and praising at intervals for attentive behavior in the other lesson. Variation in instructional strategies would be a more complex approach to planning alternative lessons, e.g., using an inductive discovery strategy in one lesson and a deductive strategy in the other. (The illustrative inquiry project in the appendix has short-term objectives and includes variation in the use of a particular skill as well as variation in instructional strategies.)

2. Decide what data to collect for evaluation of the effectiveness of the two lessons.

Data on the interactive behavior of the teacher will usually be used to determine whether the planned variations in the two lessons did in fact result in observable differences. Data on the interactive behavior of pupils may be used to determine if instructional objectives were achieved. Samples of pupil work or pupil attitude inventories may also be used for this purpose. Interactive data may be collected by use of a standard category system, but usually the teacher will need to adapt such a system to highlight the differences he/she plans to create. It is also possible for the teacher to design his/her own simple category system. (The sample inquiry project in the appendix utilized the Flanders Interaction Analysis System with

a time-line display and with the addition of special subscripts. Additional required data for this project were pupils' papers to be evaluated on the basis of divergent thought production.)

3. Teach the two alternative lessons to two different groups of pupils and collect the necessary data.

Lessons may be audiotaped or videotaped for later playback and analysis of interactive behavior. Another possibility, if a simple tallying of a specific behavior is the only interactive data to be collected, is to have another teacher observe the lessons and code behavior. Involvement of colleagues in one's inquiry projects generally operates to increase the supportiveness of the climate. It is recommended that teachers try to get feedback from their peers at one or more stages of the inquiry process. (For the sample inquiry project in the appendix, the lessons were videotaped for later coding.)

4. Compare and analyze the two alternative lessons, and evaluate their effectiveness in attaining the desired objective(s).

One criterion for assessment of teacher competence is that the teacher can demonstrate planned variation in behavior. Another criterion is that the teacher can achieve the instructional objective in at least one of the two alternative lessons. A third criterion may well be that the teacher can plan, conduct, and learn from a project in which he studies the effects of his own behavior on pupil outcomes. (In the illustrative inquiry project in the appendix the teacher was able to demonstrate behavioral differences. Samples of pupils' work are not available for us to judge whether the instructional objective was achieved.)

5. Continue to test additional alternatives or work to improve skill in using the more effective procedure.

One result of the inquiry project approach may be that the teacher generates additional variations to improve the attainment of the original objective, or other similar objectives. This could lead to improvement of instruction as the teacher identifies more and more effective alternatives. Another possible result of the inquiry project is that the teacher finds a particular skill or strategy to be slightly more effective, and decides to improve his ability to use it. This could lead to improvement of instruction as the teacher engages in further skill training. (The teacher who conducted the sample inquiry project in the appendix decided to learn more about teaching strategies that encouraged divergent thinking and began studying the synectics model presented by Joyce and Weil.)⁷

Anticipated Outcomes

The inquiry approach outlined above has not been thoroughly tested as a method for assessing teacher competence. Materials for training teachers to use the method are currently under development⁸ and will soon be field tested.

In preliminary field tests there has been some indication that the procedure of planning, teaching, and comparing two alternative lessons for purposes of assessment operates to provide a more supportive climate for change. Almost all teachers following this procedure are able to exhibit some observable differences in behavior and to identify differences in the effectiveness of the lessons. In most instances they judge one lesson to be more effective than the other, even though neither was totally effective in achieving the stated objective. Thus the comparative method is expected to result in some feelings of positive accomplishment and to provide some of the self-confidence necessary for exploration, assessment, and change of behavior.

The inquiry approach can serve to individualize the criteria on which the assessment of teacher competence is based. A teacher engaging in an inquiry project can identify the specific criteria for assessment in relation to particular problems of the class he/she is teaching. For this reason, the teacher will have more reason to view assessment as a process that contributes to improvement of instruction in his/her classroom, yielding useful knowledge about what is effective for the individual teacher in a specific setting with particular pupils.

Teachers who have learned the inquiry approach will have developed a strategy for continued self-assessment. They need not wait for an outside assessment to determine how to improve their instruction. Furthermore, teachers using the inquiry approach will be practicing ways of varying instruction. There is some indication that increased variation in itself could lead to a general improvement of instruction.⁹

While there is certainly no available evidence to suggest that inquiry projects will single handedly solve the problems attendant on the current demands for assessment of teacher competence, there do seem to be sound reasons for thoroughly testing the effectiveness of this approach.

Summary

Further development and refinement of the two assessment approaches presented in this paper will contribute much to the quality of teacher training. As the research base regarding teacher effectiveness expands and improves, assessment also can expand and improve. Teaching strategies, skills, characteristics, etc. that are shown to have positive effects upon student learning can be added to the testing and inquiry packages. Ultimately, as much more information is available regarding teacher effectiveness, as the psychometrics of measuring teacher behavior are understood and resolved, and as the interface between performance based training and the in-school teaching environment becomes more compatible and complementary, we may be ready to use assessment for such purposes as entry into training and certification to teach. In the meantime, both research into teacher effectiveness and time spent in training can be improved through the use of diagnostic and inquiry assessment strategies.

FOOTNOTES

⁹ M. D. Gall, *The Problem of "Student Achievement" in Research on Teacher Education Efforts* (Paper presented at AERA, New Orleans, February, 1973).

R. W. Heath and N. A. Nelson. *The Myth of Performance Based Teacher Education* (Paper presented at AERA, New Orleans, February, 1973)

B. Rosenshine. *Teaching Behaviors and Student Achievement* (London: National Foundation for Educational Research in England and Wales, 1971)

B. Rosenshine and N. Furst. Research on Teacher Performance Criteria. in *Symposium*, ed. by R. O. Smith (Englewood Cliffs, N.J.: Prentice Hall, 1971)

W. R. Borg. The Minicourse As a Vehicle for Changing Teacher Behavior: A Three-Year Follow-Up. *Journal of Educational Psychology*, LXIII, 6 (December 1972): 572-79

B. A. Ward. *Establishing a Standard of Performance* (Paper presented at AERA, New Orleans, February, 1973)

T. J. Ouirk, R. J. Witten, and S. F. Weinberg. Review of Studies of the Concurrent and Predictive Validity of the National Teacher Examination. *Review of Educational Research*, XLIII, 1 (Winter 1973): 89-113

N. A. Flanders. *Analyzing Teaching Behavior* (Menlo Park, California: Addison-Wesley, 1970): 19-21

N. A. Flanders and G. Morine. The Assessment of Proper Control and Suitable Learning Environment. in *Mandated Evaluation of Educators*, ed. by N. L. Gage (Stanford, California: Stanford Center for Research and Development in Teaching, 1973)

G. Morine. *Entry Procedures: How Do You Get on the Board?* (Paper presented at AERA, New Orleans, 1973)

B. Joyce and M. Weil. *Models of Teaching* (Englewood Cliffs, N.J.: Prentice-Hall, 1972)

Morine. *Transforming the Product* (Paper presented at the AERA, Chicago, 1974)

Rosenshine and Furst. Research on Teacher Performance Criteria

APPENDIX

Sample Inquiry Project

Many teachers find it difficult to get pupils to think and write creatively. This teacher decided to test two methods of stimulating pupils to think and write about the future. The specific objective of these lessons was that pupils would write a divergent prediction about "The Family of the Future." In both lessons the teacher wanted to encourage pupils to listen to each other's ideas in the belief that this would stimulate additional thinking.

Inquiry Objectives

To compare two different ways of using pupil ideas as methods of encouraging pupils to listen to each other.

To compare two instructional strategies (inductive thinking versus analogical thinking) as methods of stimulating pupil thinking and writing.

Assessment Questions

A. Was the teacher able to vary the two lessons as planned? (See interactive data collected.)

1. What differences were there between the two teaching methods planned?
2. Did these differences show up on the time line? How?
3. What differences in pupil participation were planned?
4. Did these differences show up on the time line? How?

B. Which lesson was more effective in achieving the instructional objectives?

1. What additional information (besides interactive data) did the teacher collect in order to make this judgment?

Interactive Data Collected

Lesson One: "Predicting the Future As An Historian"

Planned Procedure: 1. Ask students factual questions about their families (e.g. how many family members, who provides income) and record answers on board.

Coded Segment I	1	5	10	15	20	25	30
Accepts feeling*, uses praise							
Uses pupil idea							
Pupil initiation							
Asks questions							
Pupil response							
Lectures							
Gives directions, criticizes							
Silence							

Special Subscripts: 3-R- Teacher records pupil answer on board.
4-N- Teacher asks narrow, factual question.

* For readers unfamiliar with the Flanders 10-category system, both verbal and numerical designations are provided here.
N. A. Flanders, *Analyzing Teaching Behavior* (Menlo Park, Cal.: Addison-Wesley, 1973).

Lesson Two: "Predicting the Future As A Science Fiction Writer"

Planned Procedure 1 Ask pupils to describe a beehive Record answers on board

Coded Segment	1	5	10	15	20	25	30
1,2 Accepts feeling, uses praise				R	R		
3 Uses pupil ideas			R	R			
9 Pupil initiation			V	V	V		
4 Asks questions			B	B	B		
8 Pupil response							
5 Lectures	V	V	V				
6,7 Gives directions, criticizes							
0 Silence		6	6	6	6	6	6

Special Subscripts 3-R- Teacher records pupil answer on board

4-B- Teacher asks broad, high level question

6 - Teacher gives directions

Planned Procedure: 2 Give pupils fact sheet on families in 1890, providing similar types of information. Ask them how families in 1890 were similar to and different from families of today. Record answers on board.

	1	5	10	15	20	25	30
1.2 Accepts feeling, uses praise							
3 Uses pupil ideas					R R	P a	R R
9 Pupil initiation					V V	V V	
4 Asks questions				B B		B	
8 Pupil response							V
Lectures							
Gives directions, criticizes	6 6 6 6						
0 Silence			V V V V V V V V				

(Pupils study fact sheet)

Special Subscripts.

- 3-R. Teacher records pupil answer on board.
- 3-Pa- Teacher paraphrases pupil answer to clarify.
- 4-'3- Teacher asks broad, high level question.
- 6 . Teacher gives directions.

Planned Procedure

2. Ask pupils to draw an analogy between a family and a beehive. If the family of the future were like a beehive, what would it be like? Try to extend pupil ideas by adding own ideas to them, and ask probing questions to get students to extend their own ideas

	1	5	10	15	20	25	30
1,2							
3							
9							
4							
8							
5							
6,7							
0							

Special Subscript s.

- 3-Ex- Teacher extends pupil answer, building on pupil idea.
- 3-Pr- Teacher asks probing question to get pupil to extend own idea.
- 4-B- Teacher asks broad, high level questions.
- 2 - Teacher uses praise.

Planned Procedure: 3. Ask pupils to predict changes that might occur in families in 2020 if the trends (similarities and differences) noted from 1890 to 1970 continue. Paraphrase their answers to clarify ideas expressed. Make writing assignment.

Special Subscripts: 3-Pa- Teacher paraphrases pupil answer to clarify.

9-W- Pupils write answers at seats.

2 - Teacher uses praise.

Lesson Two (concluded) . . . Science Fiction Writer"

Planned Procedure: 3. Make writing assignment.

Coded Segment III	1	5	10	15	20	25	30
Accepts feeling, uses praise							
Uses pupil ideas							
Pupil initiation		W	W	W	W		
Asks questions							
Pupil response							
Lectures	V	V	V	V	V	V	V
Gives directions, criticizes		6	6	6	6	6	6
Silence							
	1,2						
	3						
	9						
	4						
	8						
	5						
	6,7						
	0						

BEST COPY AVAILABLE

Special Subscripts: 9-W- Pupils write answers at seats.

10 Teacher Competency Research

Barak Rosenshine

The purposes of this paper are to review current research on teaching skills and to suggest procedures for future research which can take place as part of competency based teacher education.

There are two action suggestions within this paper. The first is the suggestion that each institution engaged in competency based teacher education, conducts one experimental study per year involving teacher skills and student achievement. The second suggestion is that each institution produces one interpretative review on a teaching skill per year. Details about how this might be done and possible topics for study are contained within this paper.

State of Research on Teaching Competencies and Pupil Achievement

Most papers on teacher education contain the embarrassing recognition that the present scientific base for teaching and teacher education is primitive. That is, the number of studies which have looked at both teacher behavior and student outcomes is embarrassingly small. A diligent search will uncover less than a hundred studies. The quality of many of these studies is questionable. The results are most useful for suggesting future research; the results of these studies are *not* sufficiently strong or clear to direct teacher training practices or certification or evaluation of teachers.

In contrast, there is no lack of research on *how to train* teachers to introduce a lesson, ask questions at higher cognitive levels, redirect pupil responses, praise students, or encourage student talk. Nor do we lack for

lists of teaching skills. What is lacking, however, is research demonstrating an experimental or correlational connection between these skills and measures of change in pupil achievement and affect.

Statements on the lack of research may be surprising to many readers. Readers are invited to check these claims for themselves. Consult a textbook in educational psychology, teaching, or subject area methods and see how much research exists for oft-proclaimed teaching procedures such as matching teacher style and pupil style, asking questions, using behavioral objectives, promoting student talk, and motivating students. In these areas, as in others, I find only a few studies which took place in classrooms, and these have yielded equivocal results.

I believe that there are insufficient studies to date, to permit an evaluation of the promise of this research. Currently, an important task is to generate more and improved research so that the promise of this research and directions for future study can be evaluated more adequately.

Research Findings to Date

There have been very few attempts to summarize the current state of our knowledge on teaching competencies. In one review, the correlational research on some 25 teaching competencies is presented.¹ Nine of these 25 variables were selected as the most promising for future research on the basis of results obtained to date. A subsequent review supports, in general, the Rosenshine-Furst findings and conclusions.² It should be noted that these nine "most promising" variables were selected from a review of the correlational literature, and future experimental studies may not support all of these correlational results.

The variables below were listed in the 1973 summary, and more details are available in the 1971 volume. Those nine variables were:

1. Clarity of teacher's presentation
2. Variety of teacher-initiated activities
3. Enthusiasm of teacher
4. Teacher emphasis on learning and achievement
5. Avoidance of extreme criticism
6. Positive responses to students
7. Student opportunity to learn criterion material
8. Use of structuring comments by teacher
9. Use of multiple levels of questions or cognitive discourse

These nine variables are promising suggestions for future research. They *cannot* be used as checkpoints to assess teacher competency because the research to date is too incomplete.

At first glance, the above list of most promising findings may appear to represent educational platitudes. Their value can be appreciated, however, only when they are compared to the variables which have *not* shown signifi-

cant or consistent relationships with achievement to date. These variables are:

1. Non-verbal approval
2. Praise
3. Teacher warmth
4. Ratio of *all* indirect to *all* direct behaviors (I/D ratio)
5. Number of different teacher behaviors (flexibility)
6. Number of questions
7. Questions classified into two levels
8. Teacher talk
9. Student talk
10. Teacher experience
11. Teacher attitude toward subject matter

Against these above two lists, there is a much larger list of teacher competencies for which training is currently given. For most of these variables I have been unable to locate more than one or two classroom-based studies.

These variables include:

1. Probing of student responses
2. Selection of tasks to fit readiness level of pupils
3. Motivating pupils' attending behaviors
4. Establishing and maintaining rapport
5. Providing feedback
6. Conducting individual activities
7. Presenting information
8. Explaining subject matter
9. Giving examples
10. Using inductive strategies
11. Using deductive strategies
12. Pacing
13. Responding to student answers
14. Preventing discipline problems
15. Giving homework
16. Solving problems
17. Developing pupil self-concept
18. Developing learning to learn skills
19. Developing acceptance of responsibility
20. Selecting instructional strategies
21. Involving students in self-evaluation

The state of the art, then, is one of a lack of research. We simply do not have enough studies about the relationship between teaching competencies and pupil achievement.

Teacher Questions

The situation is more complex than merely a lack of research. There is also a need to integrate research findings into some meaning for teacher educators. At times this is a difficult task because of the scarcity of studies and the difficulty of comparing studies. These difficulties are illustrated in the mini review on teacher questions presented below.

Training preservice teachers to ask a variety of types of questions is part of any teacher training program. In some programs, instruction on types of questions is given through assigned readings or through worksheets and examples. In other programs, trainees receive additional practice in a microteaching setting and receive feedback on the number of questions they asked which were factual, analytic, divergent, convergent, comprehension, synthesis, evaluative, application, comparative, inferential, or any of the bewildering variety of categories into which questions have been classified. The objective of the training, of course, is to increase the teacher's use of a variety of nonfactual questions. Tens of training studies have been conducted, and almost all show that a teacher's use of a variety of "higher cognitive level" questions increases as a result of training.

Although the results of training studies are straightforward, the results are not so clear when one inspects correlational and experimental studies involving teacher questions and student achievement. (To the best of my knowledge, interpretative reviews of teaching behaviors and other student outcomes such as student attitudes or beliefs about oneself have not been written. Such additional reviews are urgently needed, and readers are invited to consider conducting such reviews.)

Correlational Studies on Questioning

Some 15 correlational studies were found which contained classifications of teacher questions and measures of student gain. Fifteen studies is an unsatisfactorily small number when one considers that there are at least five grade level groupings (i.e., primary, intermediate, junior high, senior high, and college), at least four subject areas, and at least two types of cognitive outcome measures (factual and comprehension). This alone yields 40 separate cells—and there are only fifteen studies, and the quality of many of them is less than desired.

Fifteen correlational studies on teacher questions and student gain reported in the last 15 years of research! Without any trouble, one could easily find 50 books, chapters in books, and journal articles, which claim to teach *how* to ask questions. In the two volumes on *Resources for Performance-based Education*, there are 26 modules on how to ask questions—and an equal number of filmstrips, movies, audiotapes, and multimedia kits.³ We are instruction rich and data poor.

Differences in definitions. Even with only 15 studies, there are problems in interpretation because variables with the same title (e.g., factual questions) are defined and coded differently in different studies. Although investigators usually classify questions into two types, many questions

which are classified as "higher level questions" (or open or interpretative questions) by one investigator have been coded as "lower order questions" by another investigator. For example, questions which require a pupil to explain how he got an answer, or questions which require the answer to a thought problem, have been classified as factual by one investigator and as higher level questions by another.

Even if one knew the precise definitions each investigator used to classify questions, we still would *not* know how many explaining questions, for example, were classified among the higher level questions by one investigator and among the lower level by another. Because of these problems, any comments on the overall results must be treated with caution.

Results of correlational studies Of the six studies in primary grades in which questions were broken into two types, significant results favoring use of higher level questions were *not* obtained in any study. These results held for reading, mathematics, science, standardized tests, and author-developed tests. In one study, open questions with a nonacademic focus were significantly *negatively* related to math gain in the first grade (but not in the third grade for similar lower class pupils). In another study, questions which asked children to give a personal preference or respond from personal experience were negatively related to gain in reading, particularly for middle class children. (In the same study, these results did not hold for mathematics gain.) Thus, the overall results for open questions in the primary grades are either nonsignificant or negative. In four of the above studies there was no significant relationship between the frequency of factual questions and student achievement. In the fifth study there was a positive, significant relationship. In the sixth study the result was positive and significant for *academic factual* questions; the correlations were negative and significant for nonacademic factual questions. Thus, the overall results for factual questions are nonsignificant but with a trend toward *positive* relationships.

with a trend toward *positive* relationships.

Overall, the four studies in the intermediate grades are more difficult to interpret than those in the primary grades because of greater differences in coding and statistical procedures. As in the primary grades, there is a trend for factual questions to be *positively* related to achievement gain, and for open questions which concern student personal experiences to be *negatively* related. In addition, one study yielded a significant positive relationship between "interpretative" questions and student achievement, but these results were not replicated in the other three studies. Indeed, in another study, there was a moderate number of divergent to convergent questions which were significant.⁴ Of the remaining two studies, one yielded all nonsignificant results,⁵ and one found a negative trend between questions regarding student interests and opinions and student achievement.⁶ This last finding appears similar to that obtained in two of the studies conducted in the primary grades.⁷

In two studies in junior high school grades, both in science, each used a ratio of higher level to lower level questions. In one, a *moderate* ratio was most functional, and this finding appears similar to that obtained in intermediate grade vocabulary.⁸ However, both sets of results must be seen as

tentative because students were the statistical unit. In the other study, a *high* ratio was most functional, but this finding was limited to the high achieving children in the sample.¹⁰ The high ratio was also most functional in the single senior high study, one conducted within social sciences.¹¹

The final study, in college social science, replicates results which appeared in studies in the primary and intermediate grades.¹² Interpretation questions were positively related to achievement; opinion questions were negatively related.

Summary of Correlational Studies on Teacher Questions

The conclusions listed below are tentative. They are based on 16 diverse studies, some of which do not have the best design characteristics.

1. The frequency of factual questions was either positively related or unrelated to achievement gain.
2. The frequency of "higher level" questions, as a whole, was usually unrelated to student gain. This finding held regardless of the outcome measure
3. In two studies which used ratios, a moderate ratio of high level to factual questions was most functional. However, in two other studies which also used ratios, it was the high ratio of high level to factual questions which was most functional. Because of this confusion, there is almost nothing which can be said about the appropriate mixture between high level and factual questions.
4. Questions on personal opinions or personal experiences, or other forms of nonacademic, open-ended questions were, when studied, consistently negatively related to achievement. This finding was obtained in primary grades and in college.
5. When questions were divided into finer groupings than high level and lower level questions, then a count of the frequency of interpretative questions yielded positive and significant results in two studies. However, the single experimental study which investigated interpretative questions did not obtain significant results.¹³

Of course, it is possible that the overall lack of results may be due to the primitive way of grouping questions. Using such gross categories as high level or lower level may not be an appropriate research approach. More studies will be necessary before one can comment on this problem.

Experimental Studies on Teacher Questions

Although correlational studies are good for suggesting experimental studies, experimental studies are the clearest way of validating ideas.

Given the amount of testimony on the importance of higher level questions, one would expect that if teachers were to increase their use of evaluative or synthesis questions, for example, there would be a corresponding increase in a class' ability to answer posttest questions involving these types of questions. The experimental studies to date, unfortunately, do not support such "common sense" notions. To date, experimental studies on variations

in teacher questions have usually found no significant differences.

To aid in interpretation, it is useful if experimental studies contained:

- a) Groups which were exposed to different questioning experiences
- b) Observational data on whether differences in questioning procedures actually occurred
- c) Data on outcomes measured
- d) Use of the class as the unit of statistical analysis.

Some 12 experimental studies of this type were found, ranging across different subject areas and taking place in grades 1 through 11. In three of these studies, teachers were trained to ask higher level questions, observational data confirmed that the treatments were implemented, and the outcome measures were developed by the investigators, to be sensitive to student ability to answer higher level questions.¹⁴ In all three, there were no significant differences between the scores for the higher cognitive level treatment and a control or a factual treatment. In two other studies the same design was used, but a standardized achievement test was the outcome measure.¹⁵ Again, no significant differences were obtained. Finally, in one study, the observational data indicated that the treatments did not differ and there were, understandably, no significant differences on the outcome measure.¹⁶

Significant results were obtained in two studies. In one, however, the teachers were supplied with *both* special materials and training in asking higher level questions.¹⁷ The post test contained items related to the special materials, items not covered in the control classes. In two other studies, teachers were instructed to teach either for recall or for concept mastery, and they exhibited behaviors congruent with the objective.¹⁸ Pupils in the recall treatment performed significantly better on the recall post test; no differences were found in the concept mastery test.

Finally, there were two studies in which the teachers followed a script which provided 70 percent factual questions and 30 percent higher level questions, or vice versa. The post test contained both types of questions. In one study, those in the second grade classrooms receiving the 70 percent higher level questions did significantly better on the post test;¹⁹ in the other study no significant differences were found between fifth grade classrooms.²⁰ In neither case were separate analyses made for the two types of questions.

Thus only two of the 12 experimental studies showed a significant effect for asking more higher cognitive questions, and neither of those was a clear study. In one, the teachers operated from a script, and this study was not replicated at a higher grade level.²¹ In the other, the instructional material was different for the experimental and control classrooms.²²

Despite the common sense notions, there is no evidence that asking higher level questions leads to greater mastery of higher level post test questions. Thus, there is little empirical support for the minicourses which attempt to train teachers to ask higher level questions. Of course, the studies described above can easily be criticized as being inadequate in design and conception, but such criticism will not create significant results. Perhaps mere increases in higher level questions are not sufficient. Per-

haps some strategy is involved. But the appropriate strategy needs to be researched, not asserted.

Current and Ongoing Studies

Although the number of empirical studies which have been conducted is relatively small, there are large scale studies currently under way which promise to add to our knowledge. The following seven studies are currently in various stages of initial work, data collection, or report writing.

Organization	Person in Charge
California Commission on Teaching	Fredrick McDonald (Educational Testing Service)
California Commission on Teaching (Phase I)	Fredrick McDonald (Educational Testing Service)
California Commission on Teaching (Phase II)	David Berliner, Betty Ward (Far West Laboratory)
Research and Development Center for Teacher Education, University of Texas	Jere Brophy
Institute for Development of Human Research, University of Florida	Robert Soar
Stanford Research Institute	Jane Stallings
Bureau of Educationally Handicapped (Project PRIME)	Martin Kaufman

In each of these studies, observational data is being or has been collected on classroom events and student achievement. In addition to the above, experimental studies involving teacher training modules are being conducted or are about to be conducted by:

Organization	Person in Charge
1. Division on Effective Teacher Education, Far West Laboratory for Educational Research and Development	David Berliner, Betty Ward
2. Stanford Center for Research and Development in Teaching	N. L. Gage

Of course, these are not the only studies being conducted. At least five other experimental and correlational studies were reported at the 1974 annual meeting of the American Educational Research Association.

Questions for Current and Future Researchers

When one learns that studies such as those listed above are being conducted, I would encourage workers in the area of CBTE to ask whether many of the skills currently being taught as part of teacher education are also being researched in these studies. For starters, one might consult the above lists of skills, particularly those which were labeled as promising and those labeled as currently unstudied in the classroom. (A useful task for those involved in competency based education would be to suggest how competencies such as the above list might be coded using a category or a rating instrument.)

The Need for Reviews of Research

The above lists of competencies suggests the need for reviews of research on these competencies. There is currently a lack of *interpretative* reviews of research. Most reviews, such as those which appear in a dissertation, are written to prove a point (or satisfy a committee). It is rare to find reviews which approach a topic with an open mind in order to *explore* a group of studies and *seek out* whatever generalizations might exist (and be forthright enough to admit when generalizations are not there).

Such reviews are needed, particularly on topics such as those listed in the tentative competency skills. If such reviews were written as a matter of course, the next steps for research and practice would be clearer. Interpretative reviews should be acceptable as doctoral dissertations. Such reviews make sense because most people do not conduct research after they receive their doctorates, but they do have to make decisions as to practice. For most people, experience in integrating a body of research would be more useful for their academic life than would the typical experience of conducting a single empirical study.

Thus, because of the need for these reviews and because of the importance of experience in reviewing, I would hope that more interpretative reviews would appear in colleges engaged in competency based instruction.

Conducting Research Within Teacher Education Programs

Teacher training programs, particularly those which are competency based, afford a unique opportunity for conducting research studies on teaching behaviors and student outcomes. (The term, student outcomes, is used to refer to a variety of cognitive, affective, and social outcomes which are attributable to schooling.) Half of the work necessary for such studies is already done as part of the normal training. That is, the trainees learn how to ask higher level questions, explain, listen, motivate, probe, and use student ideas and master the list of skills currently taught as teaching competencies. In most cases, it is easily demonstrated that trainees change their behavior in the course of training. It remains only to complete the study and determine whether this training also results in superior pupil achievement.

Such studies are relatively easy to conduct. As the first step, all trainees could teach the same unit to groups of pupils. One could use as teaching

units the materials on why birds sing²³ or on arithmetic.²⁴ These materials (as well as those used in the other studies summarized under Experimental Studies, below) contain both instructional materials and post tests with both factual and comprehension questions. Then, after training, teachers would teach the *same unit* to other groups of pupils. In both cases, of course, pupil achievement scores (and other measures of interest) would be obtained. The research questions are:

- a) Does the trainee's behavior change in desired directions between the first and second trial?
- b) Is pupil achievement superior after the second teacher's trial?

At present, there are only a handful of experimental studies which have data on both teacher behavior and student achievement, but there are over 200 institutions that are training preservice teachers in specific skills. If each of these institutions conducted *only one study each year*, our knowledge would take a large leap forward. Surely it is not excessive to ask each institution to conduct one such study per year, particularly when they are already engaged in extensive skill training.

Designs for experimental studies. The above design is presented below under Design One. It is Group I that was discussed above. There is, of course, the possibility that the experience of teaching the same unit twice is helping the teachers more than the training is. To answer this question, one might include a second group (Group II) which also teaches the unit twice. The research question would be whether pupil achievement on the second trial for Group II will be lower than achievement for pupils taught by teachers in Group I.

		Teach	Training	Teach
Design One	Group I	Unit A	yes	Unit A
	Group II	Unit A	no	Unit A
Design Two	Group III			Unit A straight
				Unit A showing results of training
Design Three	Group IV	Unit A	yes	Unit B (or Unit A or both)
	Group V	Unit B	yes	Unit A (or Unit B or both)
	Group VI	Unit A or Unit B	no	Unit A or Unit B

Design Two is a variation of Design One. In this case the teachers do their teaching after training. Each teacher teaches the unit twice—once using the skills they learned, and once without the skills. When this was

done, the lessons were observed to determine whether:

- a) The lessons differed in the training variables
- b) The lessons differed in other variables which are currently considered important for instruction, and
- c) There were differences in student learning and attitudes between the two types of lessons

It is important that the lessons differ in a) but not on b). The advantage of lessons in which each group acts as its own control is that the effects of experience can be controlled by randomizing which type of lesson is taught first.

A third type, Design Three, yields the best information, but can only be used when there are at least 50 teachers. In such a case, the trainees are divided into two groups; one group begins by teaching Unit A and the other by teaching Unit B. After training, each group teaches the other unit. In this way one can obtain data on two units of instruction and thus check whether the results are consistent across two instructional units.

One could add to Design Three by having trainees teach both Unit A and Unit B after training. By this procedure one would have a larger sample for analysis and could also determine whether previous experience with a unit affects student achievement and attitudes. If possible, a control group. (Group VI) would add to the interpretation of the results.

Number of Skills for an Experimental Study

Most experimental studies conducted to date test the effects of training in a single skill such as questioning, use of student ideas, or introducing a lesson. But studies need not be limited to single skills. It is possible to have preservice (or inservice) teachers receive this special training. It is possible that significant effects will be obtained more easily when training takes place in a number of skills.

Another procedure would be to run correlational type studies to relate teaching behavior and student achievement within each of the trials.

Materials for Experimental Studies

Three sets of materials are useful in conducting experimental studies:

- a) Modules or other instructional materials for training teachers
- b) Short instructional units
- c) Tests to assess student learning

Training materials. Teacher training materials are abundantly available.²⁵ The current task would be the validation of existing materials and not the creation of still more unvalidated materials.

The short instructional units are currently available from the researchers who have conducted studies using these units. These include the two week units in social studies and in mathematics used by Flanders,²⁶ the three week unit on ecology prepared by the Far West Laboratory,²⁷ the short units on mathematics and on social studies developed at Indiana University,²⁸

and a 30 minute unit on the black-billed gull.²⁹

The proposed studies would occur more easily if someone assembled these units and made them available to others. Such a task would be a useful contribution to CBTE assessment.

Popham and his associates have developed a series of 15-minute selections and post tests which contain selections dealing in all subject areas.³⁰ Whether 15-minute units or longer units are the most appropriate vehicles for research remains to be determined. (In studies to date the results were similar across different lengths of instruction.) Hopefully, a similar kit of materials and test instruments could be developed by assembling those which have been used in previous studies.

Tests. There are currently two organizations which are assembling and evaluating the available tests in a variety of cognitive and affective areas: the ERIC Clearinghouse in Tests and Measurement, located at Educational Testing Service, and the Center for the Study of Evaluation, located at UCLA. Each organization has published a number of evaluative descriptions of relevant instruments. In addition, handbooks of tests in specific areas are starting to appear.³¹ Thus, one who is looking for suitable tests will find that there are organizations concerned with helping them.

Teaching Competencies and Curriculum Packages

One area for teacher training and research which is currently neglected is the use of curriculum packages. By curriculum packages, I refer to packages such as Biological Sciences Curriculum Study, Elementary School Science, Taba Social Studies Curriculum, and Individually Prescribed Instruction.

These packages and others of this type include something more than the usual curriculum guide or the usual textbook. (Of course there are many which are the usual curriculum guide and admonition to "do good," but there are enough packages such as the above to merit discussion here.)

The curriculum materials packages offer at least three advantages: a) the materials, instructional procedures, and sequencing were developed by a team of specialists in content and instruction; b) the products have been revised and refined on the basis of internal evaluation and try-out feedback obtained from 20 to 200 classes; and c) a good deal of summative data is being generated which will help in conducting research on the relative effectiveness of the curriculum programs for different educational outcomes.

Although there are a large number of these programs, they can be grouped into three or four approaches. One grouping is the highly structured approach exemplified by a program such as Individually Prescribed Instruction or the DISTAR package. Another grouping might be structured discovery as illustrated by the Biological Sciences Curriculum Study programs or by Harvard Project Physics. A third grouping is tentatively called "teacher directed inquiry" and is illustrated by programs such as the Taba Social Studies program or the Holt Social Studies program. A fourth grouping is that of more informal, open approaches such as the Elementary School Science program or the model developed by the Bank Street College

of Education.

I would propose that in addition to training teachers in the lists of competencies developed by a committee, we also help teachers study these different programs and then train them in the competencies necessary to implement these different types of programs. In other words, why not bring curriculum and instruction together, instead of leaving curriculum as a junior partner to teacher training in teaching skills.

Curriculum Materials Packages Within Preservice Training

How might curriculum packages be used in preservice education? In elementary mathematics education, for example, we don't know which curriculum product, if any, the teachers will be using throughout their careers. However, as illustrated above, one can group the existing packages and the ones being developed into two, three, or four types and train teachers in the skills necessary to use these *types* of materials. Preservice teacher training can be conceived of as training teachers in the skills required by the *types* of educational products that are expected to be in use during the early part of their careers. Preservice teachers would learn the skills by using and studying the different types of curriculum packages. In elementary mathematics, for example, teachers would learn skills of informal teaching by attempting to implement the mathematics portion of the Bank Street or the EDC models; they would learn skills of structured mathematics teaching by learning to implement programs such as the Englemann-Becker or the Karnes, and they would learn skills of "modern math" by working through sections of the Greater Cleveland Math Programs or the Madison Math Program.

Curriculum Programs and Accountability

If curriculum materials packages are being used (and if development of new materials and programs is assigned to appropriate agencies), then the responsibilities are reallocated among the various groups. Teachers are responsible for implementing the program according to the instructions of the developers. If teachers decide to change the instructions, then those teachers become accountable for the success or failure of the students to learn. The developers are responsible for determining the most appropriate instructional procedures for the students whom the teachers will have, and if the teacher follows the instructions, then the developers are responsible for the outcomes. The developers and the teacher educators are responsible for providing the teacher training; the administrators are responsible for making the training available and for providing sufficient teacher training; and the public is responsible for providing the funds for the materials and the training. If the program includes parents working with the students, then the parents are responsible for their contribution.

In other words, rather than each group trying to fix the blame on the other, in this situation each group is responsible for facilitating the work of the other.

This proposed realignment of responsibilities seems analogous to that of a physician who is responsible for using the best machinery and available knowledge in his diagnosis and treatment, but is not responsible for curing heart disease and cancer. The responsibilities for finding cures and drugs belong to special and appropriate research organizations. Of course we want people cured of cancer; but will it be done by setting down a list of "performance competencies" for physicians? Of course we want more human interaction; but will it be done by setting down a list of "performance competencies" for teachers?

Needed Research on Curriculum Packages and Teaching Competencies

The above remarks on the usefulness of curriculum packages need to be tested. One needs to research whether the teaching skills which are part of the curriculum packages are indeed related to student achievement. At present, there have been only a few studies in this area. In one set of correlational and experimental studies, significant results were obtained on the importance of teacher correction procedures in a highly structured curriculum program.³² In other studies using other variables, the results have not been clearcut.³³

Summary

The major purpose of this paper is to call for more research on teacher competencies and more reviews on the results of previous studies. Because of the importance of this research, institutions involved in teacher competency training are urged to conduct one study and one review of research per year. If such work is initiated, then within two years we will have a clearer knowledge base from which to proceed in training and research. Suggestions for conducting such research are provided within the paper.

FOOTNOTES

¹ B. Rosenshine, *Teaching Behaviors and Student Achievement*, (New York: Humanities Press, 1971)

Summarized in B. Rosenshine and N. Furst, "Research on Teacher Performance Criteria," B. Q. Smith (Ed.) *Research in Teacher Education*, (Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971)

Updated in B. Rosenshine and N. Furst, "The Use of Direct Observation to Study Teaching," Chapter 5 in R. M. W. Travers, *Second Handbook of Research on Teaching*, (Chicago, Illinois: Rand McNally and Co., 1973)

² D. A. Potter and M. Lockhead-Katz, *A Critical Review of the Literature: Teacher Performance and Student Growth*, (Princeton, New Jersey: Division of Teacher Education, Educational Testing Service, 1973).

³ W. Robert Houston, et al., *Resources for Performance Based Education* (Albany, New York: State Education Department, 1973)

W. Robert Houston, et al., *Resources for Performance Based Education Supplement A* (Albany, New York: State Education Department, 1974)

⁴ G. R. Thompson and N. C. Bowers, *Fourth-Grade Achievement as Related to Creativity, Intelligence, and Teaching Styles* Paper presented to the Annual Meeting of the AERA, 1968

⁵ H. V. Perkins, "Classroom Behavior and Underachievement," *American Educational Research Journal*, (1965) 2, 1-12

⁶ R. Spaulding, *Achievement, Creativity and Self-Concept Correlates of Teacher-Pupil*

Transactions in Elementary Schools (Urbana, Illinois: The University of Illinois, U. S. Office of Education, Cooperative Research Project No. 1353)

¹ J. Stallings, et al. *Evaluation of Follow Through Models, 1972-73* (Menlo Park, California: Stanford Research Institute, 1974)

² J. Brophy and C. Egbertson, *Appendix to First Year Data of Texas Teacher Education Project*. (Austin, Texas: University of Texas Research and Development Center in Teacher Education, 1974)

³ G. R. Thompson and N. C. Bowers, *Fourth-Grade Achievement as Related to Creativity, Intelligence, and Teaching Styles*, 1968

⁴ G. Kleinman, *General Science Teachers' Questions, and Pupils' Understanding of Science*. (University Microfilms No. 65-3961, 1964)

⁵ N. F. Furst, *The Multiple Languages of the Classroom*. (Philadelphia, Pennsylvania: College of Education, Temple University, 1967)

⁶ D. Solomon, et al., "Teacher Behavior and Student Learning," *Journal of Educational Psychology*, 1964, Vol. 55, 23-30

⁷ Greg Millett, *Comparison of Four Teacher Training Procedures in Achieving Teacher and Pupil Translation Behaviors in Secondary School Social Studies*. Paper presented to the American Educational Research Association, 1969 (Austin, Texas: University of Texas, ED027256)

⁸ G. L. Miller, "Collaborative Teaching and Pupil Thinking," *Journal of Teacher Education*, (1966) 17, 337-358

Millett, *Comparison of Four Teacher Training Procedures in Achieving Teacher and Pupil Translation Behaviors in Secondary School Social Studies*, 1969.

⁹ V. M. Rogers and O. L. Davis, *Varying the Cognitive Level of Classroom Questions: An Analysis of Student Teachers' Questions and Pupil Achievement in Elementary Social Studies*. Paper presented to the American Educational Research Association, March, 1970 (Lexington, Kentucky: University of Kentucky)

¹⁰ W. L. Hutchinson, "Creative and Productive Thinking in the Classroom," *Journal of Creative Behavior*, (Fall, 1967), 1, 419-427

¹¹ R. L. Davidson, "The Effects of an Interaction Analysis System on the Development of Critical Reading in Elementary School Children," *Classroom Interaction Newsletter*, (May, 1968) (Doctoral Dissertation, Denver, Colorado: University of Denver, 1967), 12-13

¹² J. L. Ciesla and J. R. Okey, *The Effects of Training Preservice Teachers to Use Bloom's Mastery Teaching Strategy*. (Bloomington, Indiana: School of Education, Indiana University, 1974)

¹³ Willavene Wolf, et al., *Critical Reading Ability of Elementary School Children*. U. S. Office of Education, Project No. 5-1040 (Columbus, Ohio, 1967)

¹⁴ W. W. Lynch, C. Ames, C. Barger, W. Frazer, S. Hillman, and S. Wisehart, *Effects of Teachers' Cognitive Demand Styles on Pupil Learning*, 1973

¹⁵ L. J. Buggiey, *A Study on the Relationship of Classroom Questions and Social Studies Achievement of Second Grade Children*, Doctoral Dissertation (Seattle, Washington: University of Washington, 1971) (UM 71-28, 385)

¹⁶ T. V. Savage, *A Study of the Relationship of Classroom Questions and Social Studies Achievement of Fifth-Grade Children* (University Microfilms, UM 72-78, 661, 1972)

¹⁷ L. J. Buggiey, *A Study of the Relationship of Classroom Questions and Social Studies Achievement of Second Grade Children*, 1971

¹⁸ T. V. Savage, *A Study of the Relationship of the Classroom Questions and Social Studies Achievement of Fifth-Grade Children*. (University Microfilms, UM 72-78, 66, 1972)

¹⁹ Wolf and others, *Critical Reading Ability of Elementary School Children*, 1967

²⁰ Lynch, Ames, Barger, Frazer, Hillman, and Wisehart, *Effects of Teachers' Cognitive Demand Styles on Pupil Learning*, 1973

²¹ Ciesla and Okey, *The Effects of Training Preservice Teachers to use Bloom's Mastery Teaching Strategy*, 1974

²² N. R. Dodi, *The Florida Catalog of Teacher Competencies*. (Chipley, Florida: P. O. Drawer 190, 1973)

²³ W. R. Houston, et al., *Resources for Performance Based Teacher Education*. (Albany, New York: State Department of Teacher Education and Certification, 1973)

²⁴ W. R. Houston, et al., *Resources for Performance Based Education Supplement A*, (Albany, New York: State Education Department, Division of Teacher Education and Certification, 1974)

- ²⁶ N. L. Flanders. *Analyzing Teacher Behavior*. (New York: Addison-Wesley, Inc., 1970).
- ²⁷ M. Gall. Study of Teacher Higher Cognitive Level Questions. (San Francisco, California: Far West Laboratory for Educational Research and Development, 1974)
- ²⁸ Lynch, Ames, Barger, Frazer, Hillman, Wisenart. *Effects of Teachers' Cognitive Demand Styles on Pupil Learning*. 1973
- Ciesla and Okey. *The Effects of Training Preservice Teachers to use Bloom's Mastery Teaching Strategy*. 1974
- ²⁹ C. J. Wright and G. Nuthall. "Relationships Between Teacher Behaviors and Pupil Achievement in Three Elementary Science Lessons." *American Educational Research Journal*. (1970, Vol. 7), 477-493
- ³⁰ Instructional Appraisal Services. *Teaching Improvement Kits*. (Los Angeles, California 1972).
- ³¹ O. G. Johnson and J. W. Bonmarito. *Tests and Measurements in Child Development*. (San Francisco, California: Jossey-Bass, Inc., 1970)
- ³² M. A. Siegel and B. Rosenshine. *Teacher Behavior and Student Achievement in the Bererter-Engleman Follow Through Program* (Urbana, Illinois: Computer Education Research Laboratory, 1974)
- M. A. Siegel. *An Experimental Investigation of Teacher Behavior and Student Achievement in the DISTAR Instructional Program*. (Urbana, Illinois: Computer Education Research Laboratory, 1974).
- ³³ D. E. Ferritor, et al. *The Effects of Three Levels of Training on the Implementation of a Preschool Program*. Paper presented to the Annual Meeting of the American Educational Research Association, 1974
- J. Stallings, et al. *Evaluation of Follow Through Models, 1972-1973*. (Menlo Park, California: Stanford Research Institute, 1974)

BEST COPY AVAILABLE

PART III

PROGRAM EFFECTIVENESS

11 If the Program is Competency Based, How Come the Evaluation is Costing so Much?

Michael Scriven

Introduction

A competency based program is presumably thoroughly developed by the time it is in full-scale operation, which presumably means that it gets its graduates up to the designed level of quality and quantity of achievements. (We pass lightly over the fact that it may do this by lowering the design level in the light of realistic field trials.) Here at last, one feels—or is it just that one hopes?—is a program that knows what it does and says what it knows. Here we have laid to rest the misleading advertising of inflated phrases and vague promises that pervade traditional education catalogs.

Apart from the moral and scientific charms of the competency based approach long extolled in prose and poetry (somehow the poetry seems to creep back into the catalogs), there does seem to be at least one obvious pecuniary advantage about this innovation, namely the elimination of any need for fancy evaluation. Competency based education, the honest fellow, wears his—or is it her—credentials on its sleeve. They need no embellishment.

I wish I could add to your enjoyment of this evening's banquet by encouraging you to believe that the competency based path is indeed rewarded in this way, but I propose instead to contribute to indigestion by stressing that the direct cost and quality of evaluation in competency based programs that I know about are not even in the ballpark for adequacy. Not, I hasten to add, that this puts it in a poor position compared to most educational programs. While it is true that the competency based approach offers the possibility of a modest saving in one dimension of evaluation, this saving is usually not realized or not realized fully; and competency based evaluation

designs too often leave unanswered several fundamental questions about the merit of the program. My chief interest this evening is to outline the basic features of a satisfactory evaluation design.

To ensure a sweet aftertaste to this rather bitter post-coffee cordial, I shall add a word or two about the costs of evaluation, directed to the thesis that the *median net cost, to the appropriate agency, of a thorough evaluation, is negative.*

The Main Problem

The main problem with the competency based 'movement', as I see it today, is not—as the sideline cynics are always saying about innovations—that it does the same old things and just gives them a different name. There really *is* a different orientation and a different methodology, as well as a different vocabulary. The point can be quite well illustrated by reading the program for this meeting. It evidences no mere rechristening activity, but an effort by many participants to focus and explain their aims more clearly; an effort that definitely affects the content of at least some of the presentations.

The main problem, however, is still something pretty simple. It is the problem whether the *total effect* of the new orientation and methodology is worth the very considerable effort it has involved, especially in light of the dubious justification of many of the competencies aimed at. The aim may well have been successful, but it's not at all clear that it was sensible.

The analogy with the behavioral objectives and criterion-referenced testing movements naturally springs to mind. As I travel around the major metropolitan school districts talking to school principals and district or state staff, I believe I have noticed a steady deterioration in the interest in what I call the *credentialing* of objectives or criteria at the expense of a *pro forma* interest in behavioral objectives or criteria as such, or perhaps an interest in them for what they contribute to *conceptualizing* the educational activity, whether planning, performing, or reporting it. (e.g., also programming of texts.)

Precisely the same problem—as is no doubt obvious to most of you—faces the competency based movement. It *began* with a no-nonsense orientation towards pragmatic teaching and evaluation, with emphasis on pay-off skills. But pretty soon we begin to find that the concept of a competency floats free of the original identification as an indubitable necessity for satisfactory job performance, and is taken to include all sorts of juggling acts that someone has taken a fancy to. These are more process competencies, and efforts to disseminate them in the absence of proof that they're important for pay-off performance represent exactly the kind of faddist, ill-grounded approach which the competency based orientation was instigated to avoid. Only an uncompromising commitment to systematic and comprehensive evaluation of each program can avoid this kind of rot from undermining what could provide the foundations of a better approach to teacher education and to many other kinds of education. I'm delighted to see symptoms of concern with this point in some speaker summaries, e.g., N. L. Gage's reference to the need for first establishing desirability in teacher behavior. It's equally conspicuous by its absence in many other summaries.

Perhaps that's just because they are summaries; we shall all find out soon.

The Basic Requirements for Good Evaluation

Let me try to provide, in checklist form and as concisely as possible, an outline of the points that must be covered in the evaluation of a competency based teacher education program. They derive from a general checklist, which was originally developed for the evaluation of educational products, but which has since been expanded to apply to programs and projects (as well as to evaluators and proposals, etc.). This presentation will avoid the illustrative but slightly irresponsible anecdotes used in the Houston talk, where we were facing all the problems of after dinner speeches, and it will omit the details of rating scales, etc., which are required in order to implement a checklist evaluation in practice. The intention is simply to play out the many dimensions which have to be explored in the course of evaluation, and which are not met by the mere waving of the magic wand on which are inscribed the words "competency based."

Need

The first question, of course, is the one that I mentioned a little earlier—the question of the justification for this particular competency, set of competencies, programs committed to teaching us this set of competencies, etc. (For convenience, I'll focus on a component of a competency based teacher education project, which is devoted to teaching a particular competency.) The problem of credentialing any successful training program is essentially the problem of hooking up its products to a needs assessment. Can we show that the need in question is a *real* need? Needs assessments are typically done in two completely spurious ways; either by using standardized tests and assuming that short falls on these tests represent need, or by doing a survey of the population that will be affected (or that is *thought* to be the population that will be affected) by the training program. The deficiencies with each should be obvious, although both are provided as examples in current textbooks devoted in whole or in part to program evaluation. There is no need in the world to be performing at norm level on any standardized test. For that matter, there isn't any need to be performing at norm level on any criterion-referenced test, *unless* the criterion can be demonstrated to concern, e.g., survival or ability to cope with further experiences that *will* be leading one towards an adequate collection of survival skills, etc. Given the differentials in rates and directions of maturation, it really wouldn't make much sense to say that somebody had a need to perform at norm level (mastery level) even on a criterion-referenced survival-connected curriculum *except* at the exit test.

The deficiencies in the usual survey approach are equally serious. People are very poor judges of what they really need (as opposed to what they want) where the issues they are passing judgment on are either technical or complex or emotional or novel. Since most of the areas where the surveys are taken concern issues that meet not just one but all of these conditions, they are staggeringly unreliable guides to need. I'm not suggesting for one moment that there is some omniscient entity around who can sim-

ply and reliably identify the "true needs" of populations of people. People may well be the best judge; but that doesn't mean that they're the best judge in the short run, when uninformed about the issue, and when torn by conflicting concerns over it. There are, of course, areas in teacher education where these problems are not serious; but if you start looking into the question of values clarification in social studies curricula, citizenship education, sex education, propaganda analysis, ecology, race relations, career choice, sex role, and a dozen other examples from the everyday activities of teachers in the school classrooms of 1974, then you can readily see that the survey approach to needs assessment is likely to be unreliable. I'm not here to offer solutions—a solution to this problem requires some specification of the exact version of the problem that is facing one at the particular moment—but only to indicate the nature and dimensions of the problem. I'll leave this point with the single thought that needs assessments of suitable degrees of reliability can indeed usually be done, but it takes a lot of work and thought to do it.

Market

What is needed is often unmarketable, and what is marketable is often not needed; once the distinction has been recognized, it's clear that we have to investigate market separately from need, or else we will be left with a training program for which the development money will have been wasted, since it cannot be disseminated to any significant number of the centers. NIE is in the process of rethinking its commitments in the direction of dissemination. This is a most welcome move, since there has been a long tradition of assuming that dissemination will take care of itself, an assumption which shows an optimistic rather than a realistic nature.

Performance Data—True Field Trials?

Do we have a training program that really can deliver the promised competencies? Often we have one that has operated very successfully in an "intensive care" situation, guided by the inventor's hands and enthusiasm, but never tested out in the field, where it's got to sink or swim by itself. Where there have been field trials, they've often been semi-subsidized, with materials or assistance provided free of charge and that welcome little feeling of being part of the in-group being supplied without any charge at all. As long as you can coast along on the Hawthorne effect, that will work splendidly; but it's a disservice to everybody to suppose that you can disregard the extent to which you owe your success to that handy little booster.

Performance Data—True Consumers?

Who have we looked at as the impacted population for our little educational package? Obviously, its primary impact (chronologically, at least) is on the teacher trainees who are going through the programs. A good evaluation doesn't rest on its laurels after looking at the first contact. After all, the justification of this little module is meant to lie, not in its beneficial effects on teachers, but in *their* beneficial effects on their students. So we normally would need to look into the question of whether we have data on the eventual student effects. The widening ripples do not cease at that point, however; do we have data on impact on future employers? What about impact

on taxpayers? What about impact on other learning activities of the first level of students, i.e., the teacher trainees? What about special demands placed on the trainers themselves—have we investigated them as an impacted audience? The real consumers are a much less elite group than most evaluations suggest. And the news from the distant shores to which the effects finally spread is often not as good as the news from the points close to the point of first impact.

Performance Data—Critical Competitors

Useful evaluation is always—implicitly or explicitly—comparative. The real question for the person who is thinking about adopting or purchasing a competency based program is whether it is better than alternatives, not whether it is better than nothing at all. Hence, the evaluator sets up, or explores ERIC, for evidence about critical competitors. The choice is not between nothing at all, but between actual competing products or programs.

So, the choice is between a competency based program and an existing program, or between the competency based program and an alternative innovative program, e.g., apprenticeship systems with on-site workshops, etc.

Now, in the vanguard of new movements, there is always considerable resistance to having a new movement compared to the old one. The avant garde of the new curricula projects of the last decade were always talking about how inappropriate it was to compare them with the old ones. The legitimate part of this point was that it would have been unfair to compare them with the old ones, using only the old ones' criteria for the standards of comparison. However, it's not at all inappropriate to compare the new and the old on standards that embrace the best points of each. It is a matter of considerable significance to discover that Japanese children outperform their American age-peers, both on the Japanese curriculum tests and on the American curriculum tests. Whatever the explanation of this is, and it isn't necessarily a sign of the failure of the American school, the evaluator is making an extremely useful contribution when he or she identifies this difference. Now you can bet your life that competency based approaches *always* have been compared to traditional approaches on at least one dimension—namely, cost. And the fact is that it's compared on a great many other dimensions too: the extent of disruption, opposition by staff, etc., etc. It's a nice, comfortable little insulation move to suggest that the new is only to be evaluated on its own terms. But it's totally unrealistic, and it's illegitimate. It took many years for economists to discover that—to their embarrassment—a competitive market for refrigerators embraced not only refrigerators, but dishwashers and electric stoves. That is, the rational consumer quite properly (and not irrationally) weighs the purchase of a new appliance against the purchase of other possible new appliances. Well, so it is with the competency based approach. It has to earn its spurs in the free-for-all of the educational marketplace, not in some sequestered cloister.

Performance Data—Long-Term

We're very inclined to suppose that the termination of treatment is the appropriate time to measure the effects of treatment; and we thereby con-

fuse logistical simplicity with methodological propriety. It's *exceedingly* inconvenient to have to segregate funds for use in long-term funds; but it's *exceedingly* unreliable to act as if somehow the results of such a follow-up will automatically be the same as the results of an evaluation at the termination of treatment.

Performance Data—Side Effects

Has there been a systematic, scientific, and sophisticated search for side effects? It has to be borne in mind that from the point of view of the *evaluator*, the goals of the program are really not a matter of great importance. They're certainly interesting for the archivist, for the developer, and for the funding agency. But what interests an evaluator is what the program actually did, judged by its congruence with the needs of the impacted population. It doesn't really matter much whether what it did was what it meant to do, or whether it failed to do what it meant to do, but did something else instead, etc., etc. It's up to the evaluator to find out exactly what it did, and, of course, this means a great deal of fishing in murky waters where no guidance is provided by reference to goals statements. The search is not altogether without its own attractions; the murky waters of goal statements, frequently inconsistent or incompatible with the goal statements of others associated with the project, or out of date, or based on factually false assumptions, etc., are by no means obviously more attractive than those wherein side effects lurk.

Performance Data—Process

It's quite wrong to use process as a reliable guide to outcome, but that doesn't mean that there aren't legitimate reasons for inspecting process in the course of evaluation. Four reasons are mentioned here:

- a) The search for injustice. Whatever the outcomes, there are certain types of procedure that are inappropriate for moral reasons, and inspection of process must be made to see whether excess cruelty, inequitable methods of grading, etc., are important.
- b) The search for joy. The second question is whether there is pleasure in the learning experience. There are other reasons for trying to make learning enjoyable, namely, that to do so is likely to increase the probability of learning taking place. But if all else fails, if learning gains are negligible and if the cost is comparable, then we might as well take account of the question of whether the babies enjoy baby-sitting of type 1 more than they do type 2. I'd say that joy is a secondary end of education (because if it's joy you want, then for most people education has some overwhelmingly critical competitors), but that doesn't mean it shouldn't be an aim.
- c) Characterization of the process. When the evaluation is done, somebody is going to have to label the package that was evaluated. Of course, it had a label when it arrived at the evaluator's doorstep, but that label is as often as not extremely misleading; and one of the tasks of the evaluator is to decide whether this indeed deserves the title of "competency based" training or not.
- d) Causation. Inspection of process often provides valuable clues to determine the responsible agent for any changes that are observed.

In particular, it may be necessary for you to inspect process in order to settle the question of whether the alleged cause was the real cause of the changes, in situations where control groups and baseline data are impossible.

Performance Data—Causation

Expanding on the last thought of the previous subsection, we have to face the fact that almost every evaluation constitutes an investigation of a causal claim, i.e., the claim that such-and-such a procedure had such-and-such (allegedly beneficial) effects. This is one of the reasons why evaluation research is by no means an easy type of research. Quite often in teacher education, the amount of learning that takes place from peers is distinctly comparable to that which occurs from the training process itself; and the same applies to "natural" learning from field experiences as opposed to the contribution of the supervisor.

Performance Data—Statistical Significance

It is scarcely necessary to mention that evaluation frequently requires some rather elaborate statistical investigations, in order to determine the true significance of an apparent effect; or indeed in order to identify any effect whatsoever.

Educational Significance

When we are in possession of reports on the first ten checkpoints, we are in a position to tie them together and come up with a conclusion about educational significance. All too often, educational significance is identified with statistical significance of the difference between the treatment group and the control group. But of course that difference may be due to the use of a very large number of subjects (which magnifies trivial effects), to invalid tests, or to a dozen other causes, none of which represent any contribution to education. Checkpoint 11 is thus a summing-up checkpoint, where the results of detailed item analysis may come in, where investigations of instrument reliability may come in, and where it is frequently necessary to refer back to one's needs assessment data, in the light of one's refined characterization of the product, to see whether in fact the product, *as it turns out to be*, really meets the need that it was supposed to meet. Where the product—as in our case—is a competency based training module, this kind of worry translates itself into worries about whether one can view the little skill that was acquired as generalizable to the many other similar skills that will be needed in the classroom.

Cost Effectiveness

We now have to look into questions of cost, and what a very complicated business that is, indeed. The use of a C.P.A. is not enough; the use of an economist is not enough; we usually have to use an economist with special expertise in the field of school economics; and a C.P.A. with special expertise in the various types of bookkeeping used by schools at primary, secondary, and tertiary levels.

When we have adequate data on the real, the opportunity, the direct and the indirect, the discounted cash flow, and other costs and analyses, then

we can combine them with the results of checkpoint 11 to yield either a cost-benefit analysis (at least) or a cost-effectiveness analysis (in most cases where we find reasonably simple dimensioning) is possible.

Continuing Support

This is a funny little item, as befits its number and its location on the checklist. But it's something of the sting in the tail of the checklist, too. It concerns the question of avoiding orphans; when we start looking at orphans in the educational marketplace, we are much struck by the fact that the little entities that everybody wanted to be heir apparents to on one day are often entirely orphaned the next. Faddism is the great enemy of educational progress; in order to avoid faddism, we need, in the course of a serious evaluation, to look at the question of sustained support and improvement in a given educational program, project, or product. Will Sony stop manufacturing video cassette recorders in the near future, just after we've purchased a set of them for the school? Will the behavioral objectives movement indeed be steadily amplifying the libraries of instructional objectives after we convert to a commitment to behavioral objectives? Will there be programmed texts that the students can use once they're through this set of programmed texts, or will they be programmed orphans? Will the competency based movement follow through? The evaluator has to investigate such questions.

The Specific Problem of Teacher Evaluation

I will simply mention the fact that, apart from the problem of evaluating competency based teacher education, there is also the problem of evaluating the products of competency based teacher education programs, namely, competency trained teachers. Here we run into a number of other quite different problems for the evaluator. How do you get from a conclusion about the merit for the impacted population of having competency based teachers to the merit of an individual teacher? There is no easy path. Even such simple questions as how one gets from competency-mastery judgments (which one makes in the course of supervising the trainees) to competency-exercise conclusions (which is what one needs in order to be sure that his competencies will be used in the classroom)? James Popham has clearly stressed the important distinction between competency based approaches to program evaluation and competency based approaches to individual teacher evaluation; given the present instability of data on competency of teachers as individuals, we by no means have to throw up our hands about the possibility of evaluating training programs using competency based tests. Again, there is the question of whether we are going to evaluate teachers for minimum competency, or for purposes of developing a competency profile on which one could rely for, e.g., placing a specialist. The former is in much better shape, as the evaluation art stands, than the latter.

In short, while the competency based approach offers us considerable *opportunities* for improving the present state of teacher evaluation, this mainly shows that the present state of teacher evaluation is in very bad shape, not that the competency based approach is without its own problems.

Cost-free Evaluation

I promised to add a little touch of good news to the end of this list of worries, and I wish there was time to develop the good news at comparable length! But I can at least explain the general thesis that I want to propose, even if I can't get into details about the exact reasons for believing it to be true.

The general thesis is that the costing or evaluation itself is usually incorrectly done. It is usually seen as a cost on the books of either an agency or a project or perhaps a school district; the theory there reflects the fact that somebody has to make out a check to pay the evaluator. What doesn't show up there is the hidden side of the ledger, namely, the returns from the evaluation. Exactly the same point can be made about a fire department. If you ask the mayor of a city what the cost of the fire department is, then he or she is likely to reply by mentioning the figure that represents the amount of money paid over by the City Treasurer into the fire department account each year. Of course, the fire department isn't *just* a cost. The department's existence is what keeps insurance rates down in the city. The department's existence is what keeps the death rate and property damage rate down in the city. The returns from those activities of the department have to be taken into account in a correct set of social books. The same is true of evaluation. Evaluation is not a tack-on luxury, stuffed down our throats by external agencies. It is what it takes to determine the quality of one's own work. Without effective evaluation, there can be no knowledge that what has been done has been worthwhile. If, therefore, one's commitment is to worthwhile activity, then one is necessarily committed to substantial evaluation. Now, if we start drawing up a set of social books, then we have to recognize that evaluation is frequently responsible for terminating projects that are not worthwhile; and it is sometimes responsible for increasing the amounts of resources that are put into other projects of exceptional promise, thereby getting them beyond the threshold of very large scale payoffs. Evaluation, in short, should not represent a net cost to a community, such as the teacher educators. I won't say that bad evaluation doesn't cost more than it should; by my standards, that's an excellent ground for saying that it's bad, since it fails to outperform a critical competitor. I only say that a good evaluation is not only a good investment but the only means whereby you can recognize whether your investment is good or not. So, if I have presented a rather depressing list of activities that are involved in satisfactory evaluation and that are not taken care of by the move to competency based approaches, I have at least provided some kind of a sugar coating for the pill that I am suggesting one has to swallow!

Conclusion

Competency based evaluation is a means to the ends of education. Whether it becomes one more semantic tombstone in the cemetery of educational fads depends entirely on two issues; first, whether we can keep our eye on quality and quality control, and *meet* the standards of quality, not mere popularity; second, whether we can act like rational users or consumers and *adapt* on the basis of merit.

12 Task Force on Evaluating Program Effectiveness

**David Potter
Horace Aubertine**

Competency based teacher education is an operational reality at relatively few institutions. At most other teacher education institutions, the competency based movement is an emergent phenomenon that is proceeding at various rates. Evaluation of such emergent competency based programs must, of course, fulfill all of the usual functions of program evaluation: it must provide objective, impartial measurement of the effects of the program. In addition, program evaluation should form the core of what is in effect a management information system designed to provide program decision makers with accurate, objective, timely information concerning the relative effectiveness of all elements of the system. Such evaluative feedback is essential for effective management; without it, crucial decisions often will be made on the basis of hunch or expediency.

Ideal program evaluation as seen by the Task Force members measures program success against the 13 criteria discussed by Scriven during this conference. These criteria rest on what is essentially a systems approach to evaluation—an approach which considers not only the extent to which the organization is successful in reaching its stated goals, but which also considers things which are in a sense subordinate goals (for example, the level of morale among people involved in the program). Program success is seen as a complex, multidimensional concept. The evaluator should study not only the degree to which the program achieves its goal of improving education by improving the preparation of educational personnel, but also the processes through which the organization functions.

At the most basic level, the keystone of performance based education

programs is the willingness to be held accountable for achieving stated program objectives through the use of training and assessment procedures which are also open for public inspection. The openness to summative evaluation implied by this may be at once performance based education's greatest strength and its greatest weakness. The strength lies in the willingness to be evaluated on the basis of publicly stated objectives, but it is also a weakness in that premature summative evaluation could result in the dismissal of performance based education as a viable route to the training of better teachers before it has received a fair test. The Task Force believes that, although performance based teacher education must ultimately stand or fall on the basis of summative evaluation which compares teachers trained by performance based programs with those who graduated from other kinds of programs, such summative evaluation should not and cannot be undertaken until performance based programs have had the opportunity to mature.

This does not, however, mean that attempts at program evaluation should be postponed until PBTE programs have achieved some optimum stage of development at some time in the indeterminate future. On the contrary, there is a real and present need for systematic efforts at formative evaluation—evaluation which collects data on the success and efficiency of many aspects of program operation, data which are then fed back into the program itself in such a way as to allow program decision makers to optimally shape the course of the program. It is the position of the Task Force that performance based teacher education programs should be judged today, not so much by some absolute standard of accomplishment (summative evaluation), as by the relationship between their formative evaluation systems and their decision making structures. Given the state of the art, the ideal evaluation system is one which constantly provides program management with objective, reliable data on all aspects of program operation; conversely, the ideal program manager is one who bases his decisions on data obtained from just such a formative evaluation system.

Such a formative evaluation system should collect data on several aspects of program operation. Of central importance is the evaluation of program attainments (e.g., How satisfactory are the lists of competencies developed? Have competencies and training procedures been validated against a student achievement criterion? Can teachers graduating from the program actually perform the competencies which were specified as objectives? Does the classroom behavior of program graduates differ meaningfully from that exhibited by graduates from other programs?). But program evaluation for today's world must also include such process variables as the adequacy of the management system which supports the program, the locus of program decision making, the flow of communications within the program, and other factors which together ultimately determine the success or failure of the program.

Despite the importance of process variables in program evaluation, however, the crucial question for research remains the selection of adequate criteria of program success. This question is not as simple as it might appear at first glance, for the success of a teacher education program can be judged at many different levels. We might, for example, consider a program

successful only if the students of teachers graduated from the program achieved specific levels of academic success. At the other extreme, a program might claim success on the basis of data indicating that its graduates understand the principles of effective teaching and have a good basic grasp of their own subject areas. An effective formative evaluation strategy, however, samples program success on as many levels as possible. This approach permits the evaluator to be highly specific in the delineation of the program's strengths and weaknesses.

The Task Force believes that the criterion levels defined by Turner¹ provide a valuable heuristic device for this discussion. We basically contend that the appropriate level for the evaluation of a competency based teacher education program should be Criterion Level Three; a program should be considered successful to the extent to which its graduates exhibit the competencies for which they were trained.

While Level Three should be the central focus of an evaluative strategy, other levels should also be considered. Individual instructional modules, for example, should be evaluated at Criterion Levels Four through Six; obviously, trainees should be able to demonstrate understanding of specific competencies, demonstrate the performance of these competencies, and use them in front of live students. In a sense, these lower criterion levels might be considered necessary but not sufficient conditions for effective teaching; a teacher who does not understand and cannot demonstrate a competency in a limited situation can hardly be expected to use it effectively in a real classroom situation.

Finally, while it would be unreasonable to hold teacher education programs accountable for the academic achievement of the students of teachers who graduated from the program, a comprehensive formative evaluation strategy should also include research designed to assess the relationship between the competencies covered by the program and student achievement. Program decision makers and evaluators must know how well the teachers whom they trained are doing with their own students, for without this information they cannot adequately evaluate the competencies which are the basis of their programs. They also need to know the extent to which teachers trained by their program demonstrate under actual classroom conditions the competencies which they acquired in the program.

What the Task Force is suggesting, then, is that program evaluation for today's performance based teacher education programs should possess three characteristics: (a) it should be formative rather than summative; that is, it should be used not to judge the ultimate success of the program against some absolute criterion, but rather to help program managers shape the program toward ultimate success; (b) program evaluation should have a systems orientation rather than a goal orientation—evaluation should not end with the determination of whether or not program objectives have been achieved, but should also address itself to the processes by which the system functions; and (c) program evaluation should be supplemented by program validation—in addition to determining the extent to which the system reaches its goals, a research effort should be aimed at determining the validity of those goals in terms of the relationship between the competencies developed by the program and student growth.

BEST COPY AVAILABLE

FOOTNOTE

¹ R. L. Turner, "Levels of Criteria," in B. Rosner, ed., *The Power of Competency Based Teacher Education* (Boston, Mass: Allyn and Bacon, 1972)

13 Criteria for Describing and Assessing Competency Based Programs

**W. Robert Houston, J. Bruce Burke,
Charles E. Johnson, John H. Hansen**

During the past few years competency based education programs and projects have proliferated extensively. Some closely reflect the criteria set forth by Elam.¹ Other programs claiming to be CBE appear to be only slight modifications of more conventional approaches. Surveys² of CBE practices reflect considerable activity, but the quality appears to vary greatly. Some have simply translated their old programs into the "form" of CBE, while others have diligently applied CBE principles.

But both claim to be CBE. Attempting to describe or compare results of such programs is an almost impossible task.

Beginning in 1973, the Consortium of CBE Centers³ began a project to describe the various dimensions of CBE as reflected in operating programs. Such a tool could provide the basis for activities such as:

- (1) Surveys of CBE activity
- (2) Self-assessment of intent and progress by CBE programs
- (3) Planning a document to be used by professional preparation programs
- (4) Discussion device for considering the function and value of various criteria
- (5) Research in institutional change, programmatic strategies, and organizational constructs.

The following criteria serve as the basis for this effort. While still regarded

as a "working list," it represents the third major revision and considerable debate over the past year by a wide range of persons. In its final form, to be published by the Multi-State Consortium in the fall, 1974, each criterion will be supported by a set of indicators and program descriptors.

The purpose of this is to provide another tool in the improvement of professional education programs. Feedback from readers relative to these criteria and to the finished document will be appreciated.

Criteria for Assessing the Degree to Which Professional Preparation Programs Are Competency Based

Competency Specifications

1. Competency statements are specified and revised based upon an analysis of job definition and a theoretical formulation of professional responsibilities.
2. Competency statements describe outcomes expected from the performance of profession-related functions, or those knowledges, skills, and attitudes thought to be essential to the performance of those functions.
3. Competency statements facilitate criterion-referenced assessment.
4. Competencies are treated as tentative predictors of professional effectiveness, and subjected to continual validation procedures.
5. Competencies are specified and made public prior to instruction.
6. Learners completing the CBE program demonstrate a wide range of competency profiles.

Instruction

7. The instructional program is derived from and linked to specified competencies.
8. Instruction which supports competency development is organized into units of manageable size.
9. Instruction is organized and constituted so as to accommodate learner style, sequence preference, pacing, and perceived needs.
10. Learner progress is determined by demonstrated competency.
11. The extent of learner's progress in demonstrating competencies is made known to him throughout the program.
12. Instructional specifications are reviewed and revised based on learner feedback data.

Assessment

13. Competency measures are validly related to competency statements.
14. Competency measures are specific, realistic, and sensitive to nuance.
 - 14.1 Procedures for measuring competency demonstration assure

BEST COPY AVAILABLE

quality and consistency.

- 14.2 Competency measures allow for the influence of setting variables upon performance.
15. Competency measures discriminate on the basis of standards set for competency demonstration.
16. Data provided by competency measures are manageable and useful in decision making.
17. Assessment procedures and criteria are described and made public prior to instruction.

Governance and Management

18. Statements of policy exist that dictate in broad outline the intended structure, content, operation and resource base of the program, including the teaching competencies to be demonstrated for exit from the program.
19. Formally recognized procedures and mechanisms exist for arriving at policy decisions.
 - 19.1 A formally recognized policy making (governing) body exists for the program.
 - 19.2 All institutions, agencies, organizations, and groups participating in the program are represented in policy decisions that affect the program.
 - 19.3 Policy decisions are supported by, and made after consideration of, data on program effectiveness and resources required.
20. Management functions, responsibilities, procedures, and mechanisms are clearly defined and made explicit.
 - 20.1 Management decisions reflect stated program philosophy and policy.
 - 20.2 The identified professional with responsibility for decision has authority and resources to implement the decision.
 - 20.3 Program operations are designed to model the characteristics desired of schools and classrooms in which program graduates will teach.
 - 20.3a Job definitions, staff selections, and responsibility assignments are linked to the management functions that exist.
 - 20.4 Formally recognized procedures and mechanisms exist for arriving at the various levels of program management decisions.

Staff Development

21. Program staff attempt to model the attitudes and behaviors desired of students in the program.
22. Provisions are made for staff orientation, assessment, and improvement.

23. Staff development programs are based upon and engaged in after consideration of data on staff performance.

Total Program

24. Research and dissemination activities are an integral part of the total instructional system.
- 24.1 A research strategy for the validation and revision of program components exists and is operational.
 - 24.2 A data-based management system is operational.
 - 24.3 Procedures for systematic use of available data exist.
25. Institutional flexibility is sufficient for all aspects of the program.
- 25.1 Reward structure in the institution support CBTE roles and requirements.
 - 25.2 Financial structure (monies and other resources) in the system support collaborative arrangements necessary for the program.
 - 25.3 Course, grading, and program revision procedures support the tentativeness necessary to implement the program.
26. The program is planned and operated as a totally unified, integrated system.

FOOTNOTES

¹ Stanley Elam, "Performance-Based Teacher Education. What is the State of the Art?" (Washington, D.C. AACTE, 1971)

² Allen Schmieder, "Competency-Based Education: The State of the Scene," (Washington, D.C. AACTE, 1973), pp 10-11, Susan S. Sherwin, "Performance-Based Teacher Education. Results of a Recent Survey" (Princeton, N.J. Educational Testing Service, 1973); Donald W. McCurdy, "Status Study of Competency Based Teacher Education Programs in Science," (Paper presented at the Association for the Education of Teachers in Science, March 15, 1974).

³ Syracuse University—James Collins; Oregon—H. Del Schalock; Michigan State University—J. Bruce Burke; University of Georgia—Gilbert Shearron and Charles E. Johnson; Florida State University—Norman Dodi; Columbia University—Teachers College—Bruce Joyce; University of Wisconsin—M. Vere DeVault; University of Toledo—George E. Dickson; University of Houston—James Cooper, Wilford Weber, and W. Robert Houston; James Stettensen and Allen Schmieder represent USOE and John Hansen is Executive Secretary.

BEST COPY AVAILABLE

PART IV

EFFECTIVENESS OF INSTRUCTIONAL STRATEGIES

14 Evaluating Ways to Help Teachers to Behave Desirably

N. L. Gage

In its simplest form, the logic of teacher education depends on two kinds of relationships. In the first kind, the behaviors and characteristics of teachers are the independent variables, and measures of student achievement of educational objectives are the dependent variables. After causal relationships of this kind have been established, we know what kinds of teacher behavior and characteristics are desirable, in the sense that they facilitate student achievement. Then we are logically ready to go on to the second kind of relationship—that between teacher education programs and procedures, serving as the independent variables, and teacher behaviors and characteristics, serving as the dependent variables.

Logically, the first kind of relationship should be discovered and established before we can undertake to establish the second kind of relationship, simply because we do not know what kinds of teacher behaviors and characteristics we should aim at in our teacher education programs until the first kind of relationship is known. But it is obvious that we cannot afford to be completely logical in this sense. Children must be taught; hence, teachers must be educated. And neither enterprise can wait, nor has it waited over the centuries, for scientifically established evidence to appear on the kinds of teacher behaviors and characteristics that best facilitate student achievement.

The Generalizability Issue

Before dealing with the second kind of relationship—that between teach-

er education procedures and teacher behaviors or competencies—let us examine a problem that has vexed thoughtful students of the first kind of relationship. This is the problem of whether it will ever be possible to determine such relationships with enough generality for any practical value in teacher education. The issue has been around for a long time. In 1961, Bantock wrote that

. . . it is only possible to construct a model of successful teaching of a particular subject in terms of insight into the *nature* of that particular "subject" and of the particular sorts of demands that the nature of the "subject" makes in relation to the stage of development of the pupils concerned.¹

Thus Bantock was arguing that . . . not hope to find a general relationship between a teacher behavior variable and student achievement, but only a relationship that is specific to a given subject matter and a given stage of student development. Since there are many subject matters and many stages of student development, there will be many possible relationships between any particular kind of teacher behavior and student achievement. If we consider only five subject matters (for example, mathematics, science, English, social studies, and foreign languages) and four stages of student development (i.e., primary, elementary, junior high school, and senior high school), we have 20 possible combinations of subject matter and student development and thus 20 possible kinds of relationships between any particular kind of teacher behavior and any particular kind of student achievement. Consider, for example, the teacher behavior variable called frequency of "disapproval" and student achievement as measured by a test of knowledge and understanding, adjusted for preinstructional student aptitude. We cannot hope for any general relationship between these two variables if Bantock is correct. The relationship may be positive for primary grade arithmetic, negative for senior high school English, and zero for elementary school foreign languages.

This kind of argument has been made more recently by Heath and Nielson (1973) in their critique entitled "The Myth of Performance Based Teacher Education."² These authors found much fault with the volume entitled *Teaching Behaviors and Student Achievement*, in which Rosenshine (1971) reviewed more than 50 correlational studies.³ It was made even more recently, and in especially graphic form, by Maxwell (1974), from whom I cannot resist quoting at some length:

Despite the fact that proponents of PBTE readily admit that no one has ever identified the competencies of the successful teacher, little thought seems to have been given to the question of why this is indeed the case. There seems to have been remarkably little understanding of the natural science setting from which the basic model was borrowed.

The natural sciences can be rigorous because they deal with simple phenomena. They can deal in simple system models because they can safely assume homogeneity. One

BEST COPY AVAILABLE

electron can usually be assumed to be like every other electron.

But suppose that in a simple circuit each electron differed from every other, so that some wished to move from negative to positive, some from positive to negative, and some not at all, when the battery was hooked up. Further, suppose that some were pleased when a firm hand hooked up the battery while others preferred a more gentle touch—and that those who preferred the firm hand did so most strongly when firmness seemed somehow to suit the person displaying it, in the eyes of the particular electron making the judgment—and that the same individuality of judgment characterized the electrons who preferred the gentle touch.

If this were indeed the case, what would we say about the outcome of hooking up a bulb, some wire, and a battery? Obviously, we could make no simple generalization about so complex a situation.

Yet the analogy far understates the complexity of the student/teacher relationship. The truly remarkable thing is that those one presumes to be students of this phenomenon would recognize so little of its complexity. We don't have a list of the competencies of the successful teacher for the simple reason that teaching is an art—not a science. A behavior that is a valuable asset to a given teacher with a certain personality in many (but not all) situations, insofar as most (but not all) of his pupils are concerned, may well go over like a lead balloon (to most students) if affected by some other teacher with a different personality.⁴

Bantock, Heath and Nielson, and especially Maxwell are raising no trivial question. If they are right, the search for general, or even moderately general, relationships between teacher behaviors and student achievement is indeed futile. But they have raised this question on a priori grounds, as if the answer were foreordained by the sheer logic of the problem. And what they offer as a logical possibility is not necessarily an empirical fact. It was just this point that was made by Martin (1963) in a paper entitled "Can There Be Universally Applicable Criteria of Good Teaching?"⁵ Attacking the issue as an analytic philosopher, she pointed out that the fact of differences between two subject matters does not preclude the possibility of similarities in the best ways of teaching them. Whether generalizations about teaching methods across subject matters are possible is an empirical, not a logical, question or one to be settled by conceptual analysis. She concluded that Bantock's reasoning (and, by the same token, that of Heath and Nielson and of Maxwell) should not make us halt inquiry aimed at determining general principles of teaching.

Present Knowledge on What Teaching Behavior Is Desirable

Of course, that inquiry has not been stopped by these arguments. Despite

all the methodological flaws in the research reviewed and in the reviewers' methods of reviewing that research, Rosenshine (1971) and, more recently, Dunkin and Biddle (1974)⁶ have found some relationships that hold up across a variety of subject matters, grade levels, and other contexts. For example, in Rosenshine's review, of 17 studies dealing with the disapproval-achievement relationship, 12 yielded negative relationships and only 2 yielded positive relationships. The other three studies yielded mixed or doubtful results. The antigenerality argument would hold that such results should be checked to see whether they still occur when certain controls are applied. In a recent note entitled "Did You Control for _____?" Rosenshine reported what he found when he organized his review of various teacher behaviors according to type of outcome measure, grade level, subject matter, and specification of the observational variable, i.e., the operational definition of the teacher behavior. He reported that "The results were not any different from those already presented. That is, the overall results on teacher questions, for example, are unchanged if one divides the tables on questions according to outcome measures; the results on clarity or on task-orientation appear significant across all grade levels and all subject areas reported."

I am not saying that many generalizations are well-established. Relationships between teacher behavior and student achievement may still turn out, as Rosenshine is aware, to be a function of grade level, subject matter, type of achievement, urban vs. rural location, sex of student, sex of teacher, socioeconomic status of the student, ethnic variables, the student's "learning style," and so on. But the possibility of such phenomena, which the statistician would call *interactions*, does not establish their actuality. Only research can settle the issue.

In other fields of the behavioral sciences, certain relationships have held up over enormous ranges of contextual variation. For example, the relationship between scholastic aptitude and scholastic achievement has always been found to be positive in thousands of studies ranging in grade level from preschool to graduate school, from children on welfare to children in private schools, in every academic subject, in every region, and in every ethnic and sociological category. Similarly, sex differences in dominance and aggressiveness have always been found to fall in the same direction regardless of age, social class, and a host of other variables.⁷ Certain behavior modification techniques, such as those used in treating bedwetting or reducing classroom misbehavior, have been found to be effective in widely differing age levels and social classes. For example, removing attention from misbehaving children has decreased their misbehavior regularly. (See various issues of the *Journal of Applied Behavior Analysis*.)

The possibility of generalizations across subject matters deserves a bit of special attention, for it seems especially questionable. Can a good method of teaching a concept in geography also be a good method of teaching a concept in art? Can teacher behaviors conducive to understanding a principle in economics also facilitate understanding a principle in physics? For the psychologist of learning, the answer is, of course, affirmative. For him, a concept is a concept and a principle is a principle in very much a

universal form regardless of accidents of subject matter. And concepts and principles are learned, and hence can best be taught, in much the same way regardless of subject matter. Thus learning the concept of peninsula in geography is similar to learning the concept of asymmetry in art, and teaching the economic principle of supply and demand is much the same as teaching the principle of physics called Ohm's Law. At least that is the assumption guiding the search for general principles of effective teaching or teaching competencies. That search has not yet gone on long enough, nor has it failed so dismally as to justify its being called off. Serious research on teaching has only a relatively short history, as also does serious educational research of any kind. It was seldom performed on any adequate scale before about 20 years ago. Its increase in quantity is still far from adequate if we can judge by the number of studies that Rosenshine and Dunkin and Biddle were able to find for their reviews.

Furthermore, most of the studies of teacher competency have dealt with only univariate relationships. That is, only one kind of teacher behavior at a time has been correlated with a measure of student achievement. It is little wonder that the magnitude of the average correlation coefficient for any single kind of teacher behavior has been small. Indeed, we should be suspicious if the coefficients had been high because, by everything that seems plausible, a single kind of teacher behavior variable should make only a small difference in student achievement. We need studies in which patterns of two, three, and more kinds of teacher behavior are studied. Those patterns should not be studied only by means of multivariate statistics or multiple correlation coefficients. They should also be studied as sequences in the forms recently sketched by Flanders. The kinds of transition chains he described as analyzable by computer programs now available should make it possible to analyze teaching with something approaching adequate sophistication. Our counts of the frequencies of single kinds of behavior, such as disapproval, are analogous to describing a piece of music by counting the number of F-sharps it contains. We know that the melody resides not in the frequency of single notes, but in the relationships between notes. The significant dimensions of teaching may similarly be found to reside in the patterns, sequences, or chains of teacher and student behavior that occur in the classroom. If so, these new approaches to describing classroom phenomena may begin to yield more substantial explanations of the variance in achievement between classrooms when student ability is held constant.

So far, I have defended the search for the kind of relationship that will make performance based teacher education truly advantageous. I have defended the search against logical or conceptual attacks, against methodological criticisms, against premature discouragement, and against complacency about current approaches. I believe performance based teacher education should go on, even before the research successes that we seek have been won, because even without the fruits of such research it provides a tremendous heuristic advance over conventional teacher education. It exposes the questions we need to ask, and it proceeds in an optimal way on the basis of what knowledge about teacher competencies is now available.

Research on Teacher Education

But now let me turn to the second kind of relationship—that between teacher education procedures and methods, on the one hand, and teacher competencies, on the other. How can we know when a teacher education method is producing a teacher competency, or a desirable kind of teacher behavior, or the use of good models or methods of teaching, or what Gagné (1973) recently identified as more important, in his opinion, than any of these, namely, events that support learning processes? In short, how can we evaluate the instructional strategies that are used in teacher education programs?

"Knowing That" and "Knowing How"

My answer is based in part on a distinction between two kinds of competencies. This is the distinction first made, as far as I know, by Ryle (1949)¹⁰ and since then further analyzed by Roland (1961),¹¹ among others—the distinction between "knowing that" and "knowing how." The first kind of knowledge, "knowing that," refers to the ability to state factual propositions, such as knowing that Skinner coined the term "operant" or that teacher education is a controversial subject. "Knowing how," on the other hand, refers to skills or operations, and Roland added the further qualification that the distinguishing feature of such skills is that they imply learning through practice. As she put it, "Jones could not know how to swim or speak French unless he had at some time practiced swimming or tried to speak French."

My purpose in referring to these distinctions is to delimit the kinds of outcomes or objectives of teacher education with which I shall deal. In some conceptions of competency, or performance, based teacher education, all outcomes are embraced. Such conceptions include the teacher's general or liberal education, the teacher's knowledge of the subjects he will teach, the teacher's knowledge about curriculum and instruction, the teacher's knowledge about the social, historical, and philosophical foundations of education, and finally, the teacher's knowledge of how to teach.

I place most of these kinds of knowledge in Ryle's "knowing that" category. I also prefer to omit from my discussion any consideration of ways of helping teachers to acquire such knowledge and ways of evaluating techniques of giving teachers such knowledge. The reason for restricting discussion in this way is twofold. First, ways of inculcating and evaluating such knowledge are relatively well established. Teacher education's main weaknesses, in my opinion, have not been located in this realm. When it comes to giving teachers knowledge of this kind, we have done about as well or as poorly as the rest of higher education. Second, many of these kinds of knowledge are provided by parts of the university outside of schools or colleges of education, and professional educators can impose performance based approaches outside their own domains only at the risk of arousing considerable academic opposition.¹² My inclination is to try to establish the value of performance-based teacher education in those domains for

which professional educators are primarily responsible. And, of these, the one that provides the greatest room for improvement and the greatest opportunity is the realm of "knowledge how"—the realm of skills and habits in teaching. This is the realm in which we include what the teacher does in making decisions about how to teach and how he then uses the skills necessary to carry out those decisions. "Knowing that" is illustrated by the student's ability to state and even explain Archimedes' and Newton's laws. "Knowing how" is illustrated by the student's swimming. Although swimming is explained by Archimedes' and Newton's laws, knowing those laws is one thing, while knowing how to swim is another. The two kinds of knowledge are independent of one another. Similarly, knowing the laws of learning and being able to expound operant conditioning or cognitive theory is one thing, but even though those theories deal with what makes certain kinds of teaching possible, knowledge of those theories is not sufficient to enable one to teach.

It is this realization that has led teacher educators to include student teaching, or some other kind of practice, in their programs. Teacher educators may not have been aware of Ryle's distinction between knowing that and knowing how, or the many subtle elaborations of that distinction that have been formulated, but they have acted upon that distinction. It may be that distinction which most clearly lays the basis for our whole concern that teacher education be performance based. We want teachers to know *how* to teach, not merely to know *that* certain things about teaching are true. This realization has also led to all the attempts of the last decade to improve the effects of practice for prospective teachers—attempts in the form of microteaching, minicourses, and the hundreds of teacher training products that have been developed in recent years.

How to Evaluate Teacher Training

How should we evaluate the strategies that we develop to improve the degree to which teachers behave desirably? In dealing with this topic, we must set up hypotheses about desirable ways of teaching. Furthermore, we must set up hypotheses about strategies for getting teachers to behave in these desirable ways. Both of these kinds of hypotheses have frequently been made in the developmental work of the last few years.

Teacher Training Products

That inference seems justified by the plethora of teacher training products that have been developed, identified, and catalogued by organizations in Florida, at the University of Houston, and at Stanford University. In the Stanford catalog, which had the benefit of being based upon several predecessors developed elsewhere, more than 600 such products have been described in terms of more than 100 descriptors, each of which has provided the basis for an index of the catalog. The Stanford catalog is stored on computer tape, and information about teacher training products with any

given characteristic or set of characteristics can be readily obtained. Thus products intended to improve (a) question asking skills in (b) secondary school (c) mathematics teaching, can be identified quickly.

One descriptor in the Stanford Catalog of Teacher Training Products was "References describing field test results." Of the 657 products described in the catalog, only 87, or 13 percent, came with references to published field test results. For another 83, or 13 percent, of the products, the source catalog stated only that the product had been "field tested by the developer." For 22, or 3 percent of the products, reference was made to unpublished data or an unpublished report. When the developers were asked for information by mail, positive results were claimed for 7, or about 1 percent, of the products. For 12 of the products, the developer indicated that field testing was planned or in progress. For 77 products, no field testing had been performed, or it was indicated that the results were available only to the developer of the product. Finally, for 369, or 56 percent, of the products, no information concerning field testing was available in any form.

Thus far we have not determined what kind of information was provided in the 87 references to published field tests. Nor have we ascertained the bases for the developers' claims of positive results for seven of the products. Nonetheless, even without further investigation and analysis, the broad outlines of the field test information indicate that data are available in too few cases, in too little detail, and in a form too unavailable for evaluation by potential users or critics of the product.

The field testing of teacher training products can be expensive and difficult. In many cases, it may not seem important enough to justify the cost in view of the nature of the product and past experience with products of its kind, such as manuals or programmed instructional materials. The field testing of a product for validation can cost much more than the development of the product itself.

Yet it must be clear that the present state of affairs should not be allowed to continue much longer. Standards should be developed in this field comparable to the ethical and professional standards that have been promulgated by the American Psychological Association (1966) for psychological tests and manuals.¹⁴

Evaluative Experimentation

What form should such standards take? How should strategies for inculcating teaching competencies be evaluated? To a considerable degree, the broad outlines of answers to these questions are already available. In general, we can evaluate such products or strategies by using them as independent variables, or treatments, in experiments or quasi-experiments. Teachers or prospective teachers given training with the procedure or product should be compared with teachers or prospective teachers not given such training. Designs for such experiments were clearly set forth by Campbell and Stanley (1963)¹⁵ and have recently been treated with specific reference to the problem of evaluating teacher training products by Okey and Ciesla (1973).¹⁶ The latter exposition focused particularly on three

quasi-experimental designs because of the presumed difficulty, in research with teachers, of obtaining control groups of teachers. These three quasi-experimental designs are the time-series design, the equivalent time-samples design, and the posttest-only control group design.

These matters of design are well known. And the latter three designs have been well illustrated in their application by Okey and Ciesla to the evaluation of an actual teacher training product, namely, their own *Teaching for Mastery* self-instructional program. So I shall not deal further with these aspects of the problem of evaluating ways to get teachers to behave desirably. Let me turn instead to three aspects of the evaluation problem that are, as I see it, somewhat less well known and understood. These are (a) the use of criteria at various levels on the continuum of immediacy-ultimate, (b) the use of unobtrusive measures, and (c) the use of factorial designs.

My first point is aimed against a dogma which appeared as a reaction against the nonuse of student achievement as a criterion in studies of teaching. Until the 1950's, criteria of teaching effectiveness usually took the form of ratings by supervisors, principals, experts, or students. The bibliography by Dumas and Tiedeman annotated more than 1,000 studies, of which only a few used student achievement as the dependent variable.¹⁷ In 1954, Ackerman published a review of the few available studies that related teacher competence to pupil change,¹⁸ and he was followed in 1956 by Mitzel and Gross with a similar review.¹⁹ Since then, perhaps 100 studies, mostly correlational studies, of relationships of teacher and teaching variables to student achievement have appeared and have been reviewed in the volumes by Rosenshine and by Dunkin and Biddle (1974). More studies of this kind are needed—many more.

But my point is that these are not the only kinds of studies worth doing. Not everything desirable about teaching or teacher education needs to be justified on the basis of its relationship with student achievement. Achievement is at one end of a causal chain, or at one corner of a causal network. Many desirable factors in teaching or teacher education may be too remote from that final link in the chain to have a discernible effect upon it. They are desirable, nonetheless, if they can be shown to have good effects on links closer to the treatment, or independent variable, that we are evaluating. In medicine, not every feature of the physician's technique is validated on the basis of its correlation with patient recovery or longevity. It would be difficult to find such a correlation for every detail in methods of taking blood pressure, using an ophthalmoscope, swabbing a throat, applying a plaster cast around a broken leg, sewing a suture, or massaging a muscle. The criteria for the effectiveness of such competencies on the part of physicians are not patient longevity, or even patient recovery, but rather outcomes much closer in the causal chain to the particular competency in medical practice. Does the throat swab cover the tissue it should cover? Does the plaster cast endure long enough without being too heavy? Does the suture hold the wound together and leave a minimal scar? Does the massage restore circulation and relieve an ache?

By the same logic, we should ask of a technique of teacher education not only, and not always, whether it results in improved student achievement, retention, and even transfer of cognitive and social-emotional objectives.

We should ask first, perhaps, whether the teachers being educated with the technique have some respect and some liking for it. Unless they take some pleasure in the technique, intellectual or social pleasure, or both, the teacher training method will probably not be effective, because it will not be used. Second, we should ask whether the teacher training method has the intended effects on teacher behavior or performance, either in the interactive situation with students or in the teacher's preparation, planning, and decision making before he meets his students. If teacher attitude and teacher behavior are influenced as intended, we are justified in regarding the teacher training as validated in an important sense. We should not insist that every single procedure or product for training teachers be validated by its effects on student achievement.

Further, just as criteria come in hierarchies and not only those at the top are worth using, so settings for training teachers come in a hierarchy of realism. We can go from training animals to tutoring individual students in the laboratory to teaching groups of 5 to 10 students in a microteaching setting, and perhaps other intermediate settings, before we come to real classes in real schools. Not every product or procedure should be required to prove its worth on the basis of its effects in actual classrooms. Here again we should try to make reasonable inferences in a graduated sequence.

In short, proximate criteria of all sorts are justified. We are likely to make many errors in which we reject good methods as ineffective, if we always demand that every teacher education procedure or teacher training product result directly in improved student achievement.

Let me turn now to the use of unobtrusive measures. The term was originated by Webb and others in a 1966 volume by that title. It refers to "social science research data *not* obtained by interview or questionnaire."²⁰ Research on teaching has not been as prone as other kinds of behavioral science research to depend entirely on questionnaires and interviews. The observation of social interaction, especially in the classroom, has been more highly developed in research on teaching than in other fields of the behavioral sciences. Nonetheless there is need for additional ways of measuring the outcomes and processes of teaching and of teacher training, especially in the non-cognitive domains. We need ways that are not subject to the biases and weaknesses of tests, questionnaires, and interviews. Webb and his co-authors gave a bookful of examples of unobtrusive measures, which they classified as those depending on (a) physical traces of erosion and accretion; (b) the use of archives, including running records of various kinds, as well as episodic and private records; and (c) observation, either simple or contrived.

Of these, research on the evaluation of teacher training strategies has already used much simple observation and some contrived observation, depending on hidden hardware and controls, such as remote-controlled videotape recorders. But it has made too little use of the other kinds of unobtrusive measures. I can offer only a few examples of what I have in mind. If a helicopter flies over the playground of a multiracial school and photographs the upturned faces, the picture will yield an unobtrusive measure of racial integration in the playground. (I am indebted to Harold B. Gerard for this example.) Similarly, if the use of a minicourse is made

completely voluntary. how many teacher trainees even look into it, and how many voluntarily complete it? If a teacher education program is changed, what happens to the characteristics of those who apply for admission to it—such characteristics as prior grade-point average, sex ratio, ethnic and socioeconomic background, and scholastic aptitude? What happens to the longevity in the profession of its graduates? What happens to the incidence of broken windows and other kinds of vandalism if teachers are trained in methods of enhancing the students' sense that they control their own fate? What can we tell about a reading teacher's effectiveness from an unobtrusive count of the number of books checked out of the library by his students on their own initiative? What can we tell about the interest of a teacher's choice of topics, before and after the teacher has been given a relevant minicourse, by unobtrusively listening to students' conversations as they leave the class or ride home on the school bus? What can we tell about the effectiveness of a teacher's unit on nutrition by watching what his students buy in the school's cafeteria, store, and vending machines? My examples may be inadequate, but the point remains. We could benefit from greater ingenuity in evaluating teacher training strategies through supplementing present methods with ethically but effectively used unobtrusive measures.

Finally, I turn to the question of how to evaluate a teacher training system as a whole while also evaluating its parts. This kind of problem arises when we want to know not only whether the system or program as a whole has made a difference, but whether each of its components carries its weight and even whether the effectiveness of the whole system is greater, through some kind of interaction among the parts, than the sum of the effects of the parts. The classic solution to this problem is factorial design. For example, if there are two minicourses, A and B, there can be four treatment groups, namely teachers who take A only, those who take B only, those who take both A and B, and those who take neither. But if there are three minicourses, or other components, in a training system, the number of possible treatment groups becomes 8; with 4 components, 16, and so on. From such a design, one can learn about the so-called main effects, that is, the effects of each component by itself, and also about interaction effects, whereby two or more components together have an effect different from the mere sum of their separate effects.

For complex teacher training systems, complete factorial designs are unworkable simply because they require too many teachers. Further, they would yield much unpromising information, such as information about interaction effects that are highly unlikely to occur or too hard to interpret if they do occur. For these reasons, statisticians have invented confounded factorial designs, fractional factorial designs, and incomplete block designs. The details of these designs are beyond the scope of this paper. But research on teacher education should eventually make use of these designs, which sacrifice some of the combinations in systematic ways and some of the results on interactions, while becoming much more manageable.

I know of only one example of this kind of research. It dealt with an elemental analysis of the Keller method of instruction. Calhoun determined whether six elements of the Keller plan made a difference in student

achievement. In a complete factorial design, he would have had 2^6 , or 64, combinations or treatments. Instead, he had only six groups, and he sacrificed the opportunity to learn about many possible interactions. But he was able to conclude that each of the four principle elements of the Keller plan contributed to the overall effectiveness of the method. These elements were (a) using proctors to test students, (b) using a unit-perfection criterion over small units of material, (c) giving immediate feedback as to performance, and (d) allowing students to go at their own rate. Further, he found that the addition of lectures did not improve achievement.

The same logic and general design can be applied to the evaluation of systematic teacher training models of the kind being developed at the Stanford Center for Research and Development in Teaching. We hope to be able to evaluate in this way not only the effectiveness of the system as a whole but that of each of its components. I commend the approach to anyone concerned with the evaluation of instructional strategies in performance based teacher education.

FOOTNOTES

G. H. Bantock. Educational Research: A Criticism. *Harvard Educational Review*, XXXI, (Summer 1961), 264-280.

R. S. Heath and M. A. Nelson. *The Myth of Performance Based Teacher Education* (Berkeley, Cal.: Nomos Institute, 1973).

Burak Rosenshine. *Teaching Behaviours and Student Achievement* (London, Eng.: National Foundation for Educational Research in England and Wales, 1971).

W. D. Maxwell. PBTE: A Case of the Emperor's New Clothes. *Phi Delta Kappan*, LV, 5 (January 1974), 306-311.

J. R. Martin. Can There Be Universally Applicable Criteria of Good Teaching? *Harvard Educational Review*, XXXIII, (Fall 1963), 484-491.

Michael Dunkin and Bruce Biddle. *The Study of Teaching* (New York: Holt, Rinehart and Winston, 1974).

E. Maccoby, ed., *The Study of Sex Differences* (Stanford, Cal.: Stanford University Press, 1966).

N. A. Flanders. The Changing Bases of Performance Based Teaching. *Phi Delta Kappan*, LV, 5 (January 1974), 312-315.

R. M. Gagne. Observations of School Learning. *Educational Psychologist*, X, (1973), 112-116.

Gilbert Ryle. *The Concept of Mind* (New York: Barnes & Noble, 1949).

John Roland. On the Reduction of Knowing That to Knowing How. in *Language and Concepts in Education*, B. O. Smith and R. H. Ennis, eds. (Chicago: Rand McNally, 1961), pp. 59-71.

Ibid. p. 63.

Ellis Sandoz. CBTE: The Nays of Texas. *Phi Delta Kappan*, LV, 5 (January, 1974), 304-306.

Maxwell. PBTE: A Case of the Emperor's New Clothes.

* *Standards for Educational and Psychological Tests and Manuals* (Washington, D. C.: American Psychological Association, 1966).

D. T. Campbell and J. C. Stanley. Experimental and Quasi-Experimental Designs for Research on Teaching. in *Handbook of Research on Teaching*, N. L. Gage, ed. (Chicago: Rand McNally, 1963), pp. 171-246.

¹⁶ J. R. Okey and J. L. Ciesla, "Designs for the Evaluation of Teacher Training Materials," *AV Communication Review*, XXI, (Fall, 1973), 299-310

¹⁷ S. J. Dumas and D. V. Tiedeman, "Teacher Competence: An Annotated Bibliography," *Journal of Experimental Education*, XIX, (December, 1950), 101-218

¹⁸ W. I. Ackerman, "Teacher Competence and Pupil Change," *Harvard Educational Review*, XXIV, (Fall, 1954), 273-289

¹⁹ H. E. Mitzel and Cecily F. Gross, *A Critical Review of the Development of Pupil Growth Criteria in Studies of Teacher Effectiveness* (New York: Board of Higher Education of the City of New York, Division of Teacher Education, Office of Research and Evaluation, 1956) (Research Series 31, mimeographed)

²⁰ E. J. Webb, D. T. Campbell, R. D. Schwartz, and Lee Sechrest, *Unobtrusive Measures: Nonreactive Research in the Social Sciences* (Chicago: Rand McNally, 1966), p. 1

²¹ J. F. Calhoun, *Elemental Analysis of the Keller Method of Instruction*. Paper presented at American Psychological Association, Montreal, 1973

15 Task Force on Evaluating Instructional Strategies

**Wilford A. Weber
Ned Flanders**

This paper briefly describes the efforts of the Conference task force which was concerned with issues relating to the evaluation of instructional strategies. Specifically, the task force: (1) sought to identify the most pressing of those problems associated with the evaluation of instructional strategies; and (2) offered very tentative suggestions regarding solutions to the problems it identified. Because of the diverse nature of the task force's membership, its efforts reflect three separate but related sets of viewpoints and concerns with regard to the evaluation of instructional strategies: (1) the viewpoint of the educational researcher who is primarily concerned with establishing empirically validated relationships between specific instructional strategies and particular learner outcomes; (2) the viewpoint of the teacher educator who is primarily concerned with: (a) identifying instructional strategies which have been proved effective; (b) helping teacher education students and teachers to acquire and demonstrate those instructional strategies; and (c) utilizing assessment procedures which determine the extent to which a student has successfully demonstrated the ability to effectively use those instructional strategies; and (3) the viewpoint of the public school person—teacher or administrator—who is primarily concerned with identifying and using instructional strategies which have the demonstrated potential to effectively and efficiently facilitate pupil learning.

As might be expected, given the state of the art and the time it was allotted, the task force found the identification of problems related to the evaluation of instructional strategies to be far easier than the positing of solutions. Generally, nearly all of the solutions which were suggested consisted of recommendations that a particular problem be given careful

attention. That is, it was suggested that educational researchers should direct their efforts toward finding solutions to each of a number of problems identified as crucial to the evaluation of instructional strategies. The discussion which follows briefly describes the six major problems identified by the task force and notes the solution which was offered.

Operational Definition. The task force noted that the term instructional strategy is rather ambiguous. That is, the term lacks a commonly used, empirical, conceptual, or operational definition. It was claimed that educational researchers have been far from clear as to the phenomena which constitute an instructional strategy. On the one hand, there are those who define an instructional strategy as a rather simple teaching skill, such as the asking of probing questions. On the other hand, there are those who view an instructional strategy as a very complex set of instructional behaviors, such as individualizing instruction. This lack of clear understanding as to what behavioral phenomena constitute an instructional strategy creates a great deal of difficulty with regard to research efforts which focus on instructional strategies. Consequently, researchers seem to be rather unsure as to the unit of measurement to be employed; that is, should their research narrow in on very discrete, simple teaching behaviors or should they be concerned with more complex sequences of teaching behavior? It was suggested that presently, there appears to be a lack of empirical evidence to suggest the unit or units of measurement that are most appropriately used in studying instructional strategies. It was recognized, however, that recent attempts to identify those teaching behavior sequences and instructional strategies which constitute various models of teaching hold promise because these conceptualizations should foster operational definitions which will permit better decisions concerning questions of unit of measure. Consequently, it was the suggestion of the task force that educational researchers who are concerned with the evaluation of instructional strategies begin by focusing their attention on the identification of models of teaching which are predicted on assumed relationships between particular sets of teaching behaviors and the learning outcomes that they are expected to facilitate.

Goodness Versus Effectiveness. The task force expressed concern that the study of instructional strategies should focus on a determination of those instructional strategies which are most effective rather on those which are predetermined as "good." That is, the task force felt that the evaluation of instructional strategies should be a process which leads to an identification and validation of those strategies which are most effective, rather than a process which is intended to confirm that certain strategies are "good." In short, the task force felt that research efforts with regard to the evaluation of instructional strategies should test assumptions regarding relationships between particular teaching behaviors and particular learner outcomes. Members of the task force did not view that process as one of proving, but rather one of testing. On the other hand, the task force was quick to note that research efforts should be directed toward the study of instructional strategies assumed to be related to the learner outcomes which are deemed most desirable and meaningful.

Measurement. Problems regarding the measurement of instructional strategies were viewed as being enormous, if not overwhelming. The task

force recognized the need for better instruments with which to reliably and validly measure teaching behaviors and teaching environments. The task force also suggested that there exists a great need for better instruments for gathering reliable and valid data with regard to pupil learning outcomes. Further, it was noted that such data collection procedures needed to be more economical than is presently the case. Problems related to instrumentation were among the most serious noted by the task force, but members of the task force seemed to take the position that while it is important that educational researchers continue to work on the development of better instruments with which to measure instructional strategies, they must also be satisfied to measure as well as they are presently able to measure. That is, researchers must not be dissuaded from studying certain instructional strategies because present instrumentation is deemed inadequate. The view taken was that better measurement will only become possible if researchers commit themselves to measuring as well as they are able those variables which are deemed important. It was noted that too often important problems are left unresearched because researchers are unwilling to use instruments which do not meet the rigorous criteria usually employed by "hardnosed" psychometricians. In short, the task force took the position that one cannot find answers to problems unless one chooses to use the best available tools with which to search for such answers. Thus, the task force suggested that researchers must continue to improve measurement tools but must also use those tools presently available.

Design and Methodology. The task force recognized problems associated with research design and methodology to be crucial to the evaluation of instructional strategies. The task force listed a number of problems it saw as being crucial to this dimension. They noted in particular that there exists great difficulty in creating designs and procedures which isolate those particular instructional strategy variables which influence particular pupil learning outcomes. Likewise, they noted that there are serious problems associated with creating research designs which will allow one to study the relationships between specific teaching behaviors and various learner outcomes in a variety of instructional settings. The essential issue is the problem of isolating those particular sets of teacher attitudinal and behavioral variables--instructional strategies--which are related to particular learner outcomes. This problem is compounded by inadequacies related to instrumentation, research design, research methodology, and statistical analyses. Again, the task force took the position that such problems can only be overcome through efforts which recognize and confront these difficulties. That is, solutions to these problems are to be found in studies which--while recognizing the weaknesses inherent--nonetheless utilize such procedures as are presently available. This is not to discount the need for greater sophistication. It is simply to suggest that efforts in this regard cannot await procedures which are judged completely adequately, they must be viewed as tentative and developmental.

Statistical Analyses. Closely related to problems of measurement, design and methodology are problems associated with statistical analyses. The task force recognized that studies which are intended to examine relationships between instructional strategies and particular learner outcomes

are faced with a multiplicity of problems with regard to statistical analyses. However, this is an area in which the task force saw that a great deal of recent progress has been made. The task force expressed confidence in the ability of statisticians to find increasingly better ways in which to statistically isolate important variables. This then was one problem area in which there seemed to be a great deal of enthusiasm for present developments.

Dissemination. The final major problem identified by the task force had to do with the need to make research findings more readily available and useful to teacher educators and teachers. That is, it was felt strongly that research results are too often not available or not useful to practitioners. It was felt that research efforts had to be conducted in environments which are familiar to classroom teachers and common to normal school situations. Further, it was felt that research findings had to be put in a form that classroom teachers and teacher educators can understand and internalize. Further, these findings have to be disseminated in ways which make information readily available to practitioners, whether they are teacher educators responsible for educating teachers or are teachers seeking to increase their own effectiveness.

Summary. The task force identified six major problems related to the evaluation of instructional strategies. Each was viewed as important and as a problem needing to be solved. Few solutions were offered by the task force. Rather, it suggested that solutions were only to be found in the commitment of greater effort to what remains one of education's greatest problems.

16 Priorities and the Nature of Evidence

Allen A. Schmieder
Jorie Mark

The Problem

Leading decision makers concerned with the directions, structures, and support bases of our American society (of late being joined by more and more leaders within the educational ranks) claim that educators have little, if any, informed experience—*accumulated evidence*—upon which to base decisions about anything that is important in education. Even the teaching profession, the group that should have the best firsthand knowledge about what works and what doesn't work, is generally in favor of maintaining the status quo until more "exact evidence" about the educational process is available. Educational researchers are scrambling to prove that teaching does make a difference and that educators do know some things that are not well known to the average person on the street.

In less than 200 years, the educators of this nation have built what is probably the world's most democratic and effective educational system. Although the heavens and many other equally inspiring forces were probably on the side of this 85 billion dollar, 60 million participant enterprise, it neither grew like topsy nor evolved without considerable attention to evidence and thoughtful decisions. Consequently, it seems ludicrous to charge that today's educators have no evidence upon which to base program decisions. Because of the massive size of the educational enterprise, the continuity of its existence, and the unusual importance placed upon it by the citizens of this Nation, it is probable that educators are steeped in more evidence about the nature of their trade than are practitioners of any other single art, trade, or profession in the American social system.

The purpose of this paper is not to rally opposition to the narrow hard evidence (whatever that is) worshippers or the educational doomsayers, but to begin a dialog about how the profession can more systematically organize and use the existing National Storehouse of Education Materials and successes that have been so richly and fully built over the last 200 years

Some Expected Outcomes

Short Range

- 1 Clear statement of the problem, recommendations about how to deal with it, giving priority to the recommendations, a strategy to implement the recommendations, and a determination of the next steps to be taken
- 2 A first cut on the nature of the National Storehouse
- 3 A beginning list of locations and persons engaged in major component development with a brief description of their programs
- 4 A beginning bibliography of publications relating to the building and use of the National Storehouse
- 5 A first draft of a publication presenting the Storehouse case to leading educators
- 6 A specific strategy for further exploring the Storehouse concept

Long Range

Primary

- 1 Better definition of the levels of evidence/successes available in the Storehouse
- 2 Better knowledge of the availability of evidence/successes at these levels
- 3 More systematic relationship between evidence and program/policy decisions
- 4 More systematic design for utilizing and adding to the Storehouse
- 5 Greater use of the National Storehouse
- 6 Greater public awareness of the successes of American education
- 7 A plan for more effective sharing of successes within particular educational specialties, e.g., competency based education
- 8 A plan for more effective sharing of successes between particular educational concepts, e.g., competency based education and interaction analysis or competency based education and teaching centers
- 9 A plan for more effective sharing of successes between teacher education and other professions, e.g., competency based teacher education and competency based dental education
- 10 A plan for more effective sharing of successes between education

and business and industry regarding similar systems, e.g., storage and retrieval of airlines information, human relations vis a vis the community in education and human relations vis a vis the community in urban planning

- 11 A plan for more effective sharing of successes between education and business and industry regarding the relationship between their respective Storehouses and policy decision making

Secondary

- 1 More effective systems of storage and retrieval of validated products and processes
- 2 More effective systems of validating products and processes
- 3 Better understanding of research and development needs
- 4 Less duplication and overlap of research and development efforts
- 5 More systematic relationship between theory and practice
- 6 Development of levels of evidence regarding educational success

The Nature of the National Storehouse

What It Is

The National Storehouse is a collection of evidence—materials, programs, ideas, concepts, and practices—of what works in education today. The Storehouse exists. What does not exist is an organizational framework that could point up gaps in what we know, stimulate efforts to close them up, and provide a typology that would embrace new, as yet unthought of problems. The missing organizational framework would make the evidence easily accessible to users and would provide information on validation.

What It Is For

The National Storehouse would be a user or consumer based index with entries classified by topics focused on use or intended users. It would therefore be indexed in terms of problems or tasks an educator must cope with in order to be effective

- the practitioner identifies a teacher competency for which he wants to provide training, so turns to the National Storehouse for a search of what is available
- the policy maker is charged with formulating a new piece of legislation and needs information about it (e.g., reading program), so turns to the National Storehouse for research evidence and a bank of knowledgeable people
- the high school teacher of English wants training in how to cope with student response to literature, so enters the system seeking English, high school, student response—and the level of evidence desired. The computer search would identify the appropriate alternative materials relating to the problems as described

--- the policy maker a school board, interested in improving its elementary staff development program--- could enter the Storehouse by identifying school personnel, teacher training, elementary school, in-service, in order to see if materials exist that address this question.

In organizing the National Storehouse, therefore, we address the question, how will it assist

- teachers
- teacher educators
- legislators; national policy makers
- researchers
- educational consumers
- educational administrators
- lay administrators

What It Does

The National Storehouse identifies what we have and, as an immediate outgrowth, identifies what we need. The National Storehouse, therefore, shaped by need (or the consumer), identifies for the potential user the origins (theory, experience, hunch, invention) of the evidence; then it explains its current state of development (idea, pilot test, programmatic test, experimental test), its target, its relevance to broad categories of problems (policy decisions, research and development programs, practical application in teacher training, and organization and administration of schooling).

Organized so it can answer the question, "What do we know about X?" and open-ended as well, the National Storehouse could classify evidence in two ways:

- 1) by the categories of information a user desires;
- 2) different conceptions or categories of educational problems. This would permit the user to employ his/her own search strategy and also invite him to look at the items in a variety of different ways

Such an index could be constructed logically, based on some kind of analysis of problem areas. In a word, the system should be developed deductively from what a consensus of practitioners and laymen say we need, not inductively from what we know we have—although the latter would also be included.

Retrieval and Dissemination

A user- or needs-oriented National Storehouse must assure easy access to practitioners and policy makers.

The first step toward easy retrieval might be the organization of the index of storehouse materials into a simple directory. Contents might be divided into: (1) printed materials, (2) audio-visual materials, (3) ideas, (4) people and places. Listings should be short—product, contact name and address, one-line summary, extent of validation, cost. The directory should be set up as a loose-leaf binder so that pages may be printed and distributed for insert. It should also be free, or the cost should be nominal, covering only

mailing expenses. Obviously, then, the directory must be funded by Federal or foundation moneys. This foundation or Federal grant should be large enough to permit some subsidization of the producers of the items listed in the catalog while producers seek ways to bring down production costs.

Unless materials costs are brought down, the easy access suggested above for practitioners will not be accomplished.

But to get that directory into the hands of the practitioners will require a major dissemination program. Such a program should be launched by a communications expert who is competent to utilize all media approaches to the problem. Such approaches would include such standard methods as direct mail notifications to local and state education agencies, to associations and unions, to teacher centers, and to the various national, state, and local professional education associations. But the dissemination program should go much further. Exposure should mainly involve public service time and space in the education press (national, state, and local); the education trade and professional press, local, closed-circuit and cable radio and television; local and state conferences; displays at National conferences, and so on.

The message should be simple. American education is probably the most experienced and diverse in the world. And much of what has been learned about the way to do things better is available for educational program planners.

Costs for such a program should be explored. A followup to (but also simultaneous with) the dissemination program might be the use of trained leaders who could go out on request to work with local teachers on use of the materials and methods. The communications expert could make suggestions on how best to coordinate this. But unless state and local education agencies—or a foundation—can underwrite such a program, costs might well be prohibitive.

Quality Control

But a mechanism might be developed to inform users about the quality of the evidence.

Evidence should be thought of as a device for either hypothesis testing or decision making. Evaluation criteria differ, depending on which of these two purposes the evidence is to serve. Evidence is generally "harder" for hypothesis testing than for decision making, although this may account for bad decision making.

The second factor determining the quality of evidence is the stage of development of the idea, program, or product. Different evidence is expected for a product in its earlier stages than in its "finished" stages. Thirdly, things that work—hard evidence—within the context of an integrated and sequenced program may not work when taken out of that context.

Rather than search for "hard" evidence, the functions of rating might better be served by classifying items according to where they are in a sequence of steps in the research and development process rather than rating them. The sequence might run from armchair ideas, entirely untested, through ones tried out in laboratory experiments, in simulation or micro-

teaching settings, to those field tested and ready for dissemination. It should be possible to define a set quite objectively, and where a piece of evidence lies in this sequence should be at least as useful to know as how it was rated by somebody. If this kind of rating system was employed, it would be possible to enter the Storehouse asking for the level of development (e.g., how much the Storehouse had in that subject at a particular moment that had been thoroughly field tested). A display could be called for which would show such a quality control process by area, and it would be easy to see where weaknesses and strengths lay-- a useful tool for planning, for communicating with legislators, etc.

How To Organize It

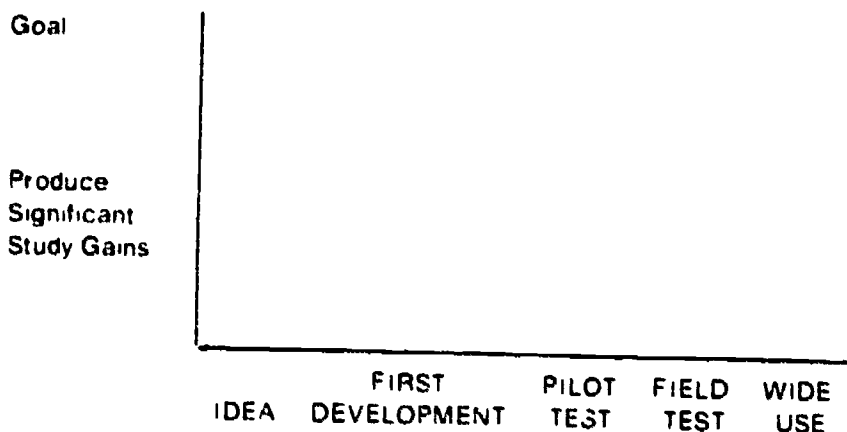
A number of models have been proposed, and they are by no means mutually exclusive.

Model A

- I. Develop information sources
 1. Gather available materials (including ideas) under four broad headings: (a) teaching skill or performance, (b) curriculum materials and design by intended outcomes, (c) training of educational personnel, (d) organization of instruction and schooling.
 2. Prepare for each item a developmental history—who developed the product or idea, its theoretical or other sources, its use, its current status.
 3. Classify each item by the criterion of its effectiveness that the user has selected and by other criteria to which it might be relevant.
 4. Gather evidence from the developer and other users about evidence relative to effectiveness with respect to the criteria.
 5. Sort items by their relevance to policy, research and development, or practice.
- II. Develop the tagging system so that anyone who wants to search the system can do so.
 1. Invite groups of potential users to ask questions of the system.
 2. Tag items by questions.
 3. Identify unanswerable questions.
 4. Determine whether the question is unanswerable because of the classification system or because there is no information in the system.
- III. Develop validation of system.
 1. Invite conceptualizers of problems to ask questions of the system (e.g., Jencks asks the system for information on school effectiveness; Coleman asks the system for information on the effects of desegregation).

2. Invite a group of people to see if they can "break" the system (e.g., ask it questions it cannot answer)

Ultimately, evidence might be categorized by eventual goal and stage of development

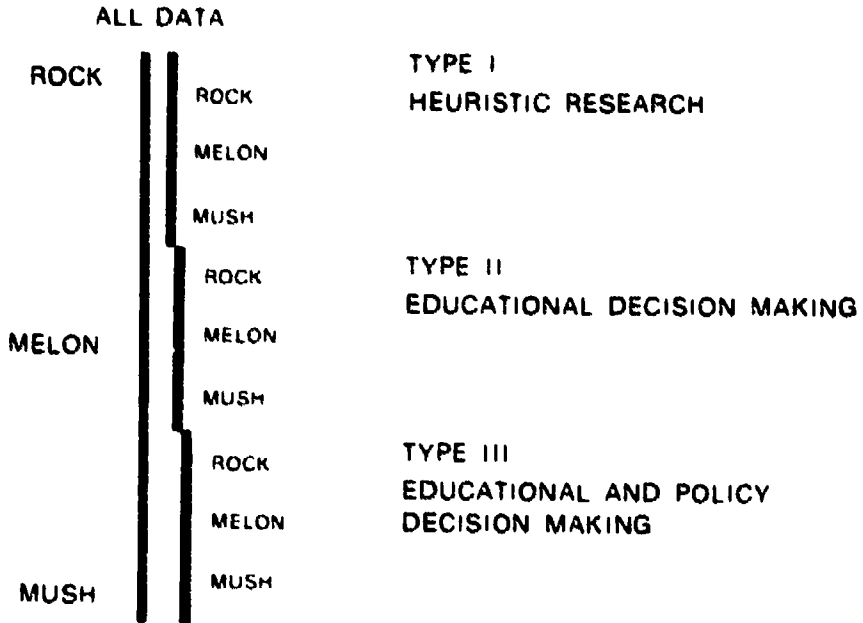


Model B

1. Develop use/user categories:

Primary User	Primary Use
Teachers	Education of Children at Various Levels
Teacher Trainers	Preservice Teacher Education
Teacher Trainers	Inservice Teacher Education
Decision Makers	Educational Decision Making at Various Levels
Decision Makers	Policy Decision Making at Various Levels
Researchers	Heuristics
Evaluators	Mission Research
General Public	Synthesis

Each Use or User category would generate its own discrete, though perhaps overlapping, set of categories designed to make sense to those interested in it for that purpose (e.g., DISTAR might be of high interest for elementary teachers, moderate interest for teacher trainers, little interest for decision makers, no interest for heuristic researchers, moderate interest for mission researchers, and be of great interest for synthesizers, and of great use in dissemination to the general public).



1 Seek answers to following questions

- 197

5. How would Storehouse be managed?
 - A. By teachers, college types, and research types?
 - B. By NEA, AACTE, AERA, ASCD, AFT, ATE, and other organizations?
 - C. Other ways?
6. What categories of organizations will allow the user to enter the Storehouse?
 - A. Category I — Levels of Evidence by Results
 1. Paper and pencil test
 2. Verbalize or write what one learned
 3. Perform in a simulated situation
 4. Perform in a classroom situation
 5. Short-term results with students
 6. Lasting (long-term) results with students
 7. Other
 - B. Category II — Subject Areas
 1. English
 2. Math
 3. Other (including Special Education and Early Childhood)
 4. Generic
 - C. Category III — Teacher Activities
 1. Planning
 2. Questioning
 3. Diagnosing
 4. Many more
 - D. Category IV — Setting
 1. Urban
 2. Rural
 3. Other
 - E. Category V — Grade Levels
 1. Primary
 2. Intermediate
 3. Other
 - F. Other Categories
7. Content of each entry
 - A. Description of material
 - B. Cost
 - C. Where available
 - D. Time for use
 - E. Source and nature of development
8. Materials should be "keyed" to other big questions of:
 - A. School finance
 - B. School policy
 - C. Social concerns
 - D. Other

Model D

CLASSROOM PRACTITIONERS				
NONCLASSROOM PRACTITIONERS				
DISSEMINATION USE				
DEVELOPMENT USE				
RESEARCH USE				
POLICY USE				
AFFECTIVE EVIDENCE				
TESTIMONIAL EVIDENCE				
STATISTICAL EVIDENCE				
HISTORICAL EVIDENCE				
NO EVIDENCE				
	A THEORY BASE	MANY THEORY BASES	SOME THEORY BASES	NO THEORY BASE

The figure illustrates a 120-cell cube which depicts a multiple category system which would be used to analyze, sort, or describe Storehouse materials.

The category system would have three dimensions--one set of five descriptors of types of evidence, one set of four descriptors of theoretical bases, and one set of descriptors of user categories. The last would be subdivided into multiple user-determined subcategories

Evidence Descriptors

The descriptors related to evidence deal with the type of evidence that supports the material. They are

Has Affective Evidence

"I like it "

"It had a positive effect on me "

"It was lousy "

Has Testimonial Evidence

"We did it and it works "

"It had no visible effect "

"Only three teachers are now using it "

Has Statistical Evidence

"A statistical random sample

" .0001 "

Has Historical (Judgmental, Philosophical, Theoretical) Evidence

"According to NIE "

"For 100 years "

Has No Evidence

" 0 "

Theoretical Categories

It is conceivable that if one could name the various theories currently supporting teacher training material, one could establish a set of descriptors accordingly. Until such is possible, a four-part category system might be usable

Has a Theoretical Base

Has Some Theoretical Base.

Has Theoretical Bases

Has No Theoretical Base

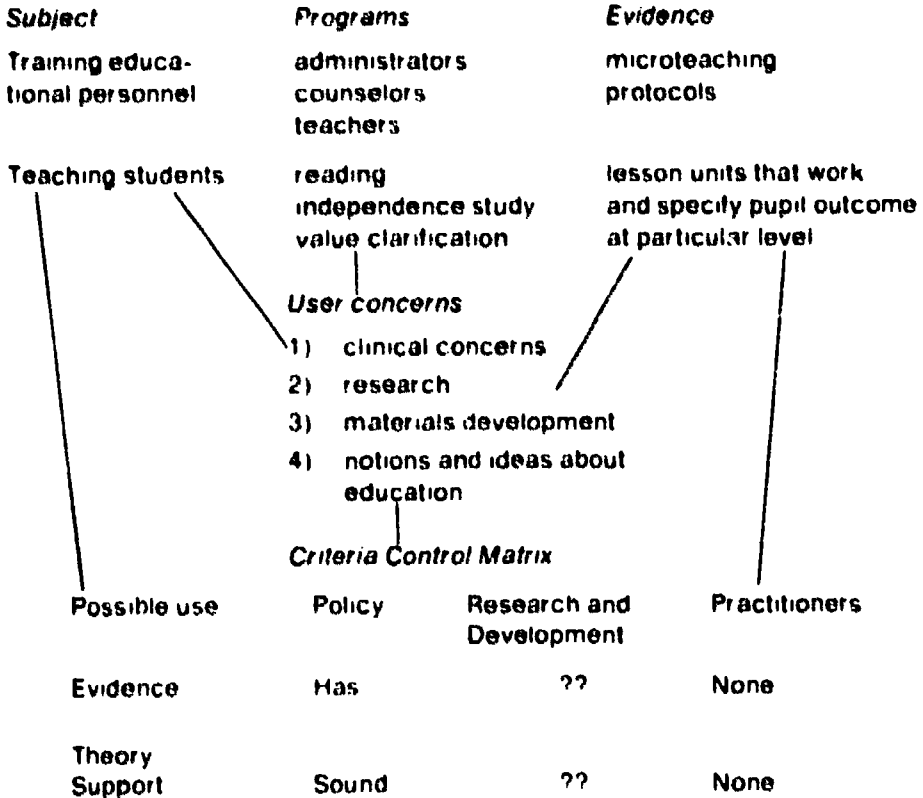
User Categories

These categories are basically undetermined. They will name as the descriptors those which will be used to extract information and material from the Storehouse. The categories will be determined by surveying of a particular group of users as to "What kind of information or material they need?" "What questions will they ask of the Storehouse?" "What ?" While any category of "User" could form a group to be surveyed, at least six can be readily identified:

1. Policy makers
2. Research users
3. Developer users
4. Dissemination users
5. Nonclassroom practitioners
6. Classroom practitioners

Model E

BEST COPY AVAILABLE



Model F

The National Storehouse of Educational Materials/Successes: Evidence Needed for Policy Making/Program Building

- Define Hard Evidence (Class I Evidence)
- Estimate Demand/Need for Hard Evidence
- Estimate Availability of Hard Evidence
- Estimate Gaps Between Demand and Availability
- Estimate Cost and Time to Fill Gaps
- Recommend Alternative Strategies for Filling Gaps
- Define Less Hard Evidence (Class II Evidence)
- Etc. (Repeat steps above until supply and demand are equal)
- The Need to Educate the Public about
 - What evidence is
 - What is available
 - What evidence professional educators will use and why, i.e., we explain and tell them, they don't lay it on us!

Some Selected Examples of National Storehouse Components

The National Storehouse of Educational Materials/Successes: Teacher Education Component—Studies of Teacher Effectiveness

- 1896 Characteristics of the best teachers as recognized by students (Kratz)
- 1905 First quantitative studies relating competence of teachers to various other characteristics (Meriam)
- 1930 An analysis of 209 different systems for rating teachers (Barr and Evans)
- 1945-46 Intensive analysis of teacher's role in classroom (Anderson, et al)
- 1948 Survey of investigations into the measurement and predictions of teacher effectiveness (Barr)
 - 79 studies analyzed
 - 46 used supervisor's or principal's ratings
 - 15 used grades in student teaching
 - 18 used measure of changes in pupils
- 1951 Assessment of socio-emotional climate in learning situations (Withall)
 - 7 categories of teacher statements
- 1954 Identification and review of all major teacher effectiveness studies completed between 1900 and 1952 (Morsh and Wilder)
 - 900 references
 - 360 abstractions
 - 20 predictors of effectiveness identified
- 1958 Analysis of 1,000 studies on teacher effectiveness (Mitzel and Cross)
 - 20 of the studies (2%) involved actual measures of teacher effectiveness
- 1960 Analysis of teacher characteristics (Ryans)
 - 6,000 teachers in 1700 schools in 450 school systems
 - 12 general "trends"
- 1966 Analysis of verbal behavior in the classroom (Bellack)
 - 4 pedagogical moves
 - 4 functionally different types of meaning communicated by teachers and students
 - 5 general roles for classroom language game
- 1971-73 Search for validated teacher characteristics or competencies (Rosenshine and Furst)
 - 50 studies
 - 11 main characteristics

- 1972 Analysis of student learning (Jencks)
Most tests used to measure student learning
insensitive to differences in teaching behavior
- 1960-74 Evaluative teaching units
Flanders -- social studies/arithmetic
Schautz and Nuthall -- electricity
Joyce -- social studies
Noy -- literature
- 1974 Exhaustive review and assessment of studies of teacher effects on students (Dunkin and Biddle)

The National Storehouse of Educational Materials: References

Teacher Training Products The Study of the Field (750). * Gage
Protocol Materials Catalog (112). Florida SEA
A General Catalog of Teaching Skills. Turner
A Catalogue of Growth in the Pedagogical Domain. Hudgins
The Houston Module Bank (225). University of Houston
NIE Research and Development Center Products (est. 100)
Far West Laboratory Minicourses (20). Berliner, Borg, Flanders, Ward
Validated Teacher Characteristics or Competencies/Process Variables
(11). Rosenshine and Furst
Critical Elements in Teaching (42). Cruickshank
Sharing Educational Successes—ESEA Title I
Sharing Educational Successes—ESEA Title III
Tested Modules in Competency Based Education (Googol). 120 CBE IHE's
Data on Teaching and Learning in Theses and Dissertations in Education
and Related Fields (Googol-plex)

* Number in parentheses indicates approximate number of products included

The National Storehouse of Educational Materials/Successes: Teacher Education Component—Some Beginning References

- Barr, A. S. "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations." *Journal of Experimental Education*, XVI (1948), 203-283
- Cooper, James E. *A Survey of Protocol Materials Evaluation*. Tampa, Florida: Leadership Training Institute in Teaching, University of South Florida, 1974
- Domas, S. J., and Tiedeman, D. V. "Teacher Competence: An Annotated Bibliography." *Journal of Experimental Education*, XIX (1950), 101-218.

- Gage, N. L. et al. *Computerized Catalogue of Teacher Training Products*. Research and Development Memorandum No. 116. Palo Alto, California: Stanford Center for Research and Development in Teaching, Stanford University, 1974.
- Huggins, Bryce B. *A Catalogue of Concepts in the Pedagogical Domain*. St. Louis, Missouri: Washington University Graduate Institute of Education, 1974.
- Guidelines for Delivery of Development Products* (Draft). Washington, D.C.: National Institute of Education, 1974.
- Protocol Catalog: Materials for Teacher Education*. Tallahassee, Florida: State Department of Education, 1974.
- Schmieder, Allen A., and Joyce, Bruce. *The National Storehouse* (for CBE), two versions - full and abridged. Washington, D.C.: USOE, Office of Career Education, Education Systems Development, 1974. (Draft)
- Schmieder, Allen A., and Mark, Jori, eds. *The National Storehouse of Educational Materials/Successes*. Washington, D.C.: USOE, Office of Career Education, Education Systems Development, 1974. (Draft)
- Sharing Education Success: A Handbook for Validation of Educational Practices*. Washington, D.C.: United States Office of Education, National Advisory Council on Supplementary Centers and Services, 1974.
- Turner, Richard L. (ed.). *A General Catalog of Teaching Skills*.

**The National Storehouse of Educational Materials/Successes:
Profile of Some Potential Experience Bases**

- 1 16,500 school districts
 - 700 community schools
 - 325 schools using individualized instruction
 - 4,500 teaching centers
 - x alternative schools
- 2 50 states
 - 20 CBE mandates
 - 9 career education mandates
 - 4 teacher center mandates
- 3 1,200 teacher training institutions
 - 120 CBE programs
 - 600 teacher centers

**The National Storehouse of Educational Materials/Successes:
Some Interesting Facts**

- 1 Many powerful educational policy makers do not think that it exists.
- 2 Many educators feel that it is already too bountiful and needs no further money for development.
- 3 Many feel that ERIC has the Storehouse well under control.

4. Storehouse materials with less validation sell faster.
5. Most new validated materials are either on a shelf somewhere or used in a single classroom or program.
6. Most consumers do not like to go to some "other" place to use catalogs and materials banks/resources—they want them located in their office.
7. The Newsletter is one of the highest demand consumer items in the dissemination business.
8. There is a very thin market for many of the most important materials developed.
9. Regarding product validation—what might be hard evidence to one kind of consumer is useless to another kind!
10. Many validated approaches are born in theses and dissertations—and eventually die there.
11. Most producers are reluctant to share materials (a) which are unvalidated; (b) which might be plagiarized or copyrighted by someone else.
12. Publishers are reluctant to produce high risk materials.
13. All producers are potential consumers and all consumers are potential producers.
14. College and university staff generally feel that they are the only ones sufficiently qualified to produce valid materials for widespread use—school staff generally feel that college and university staff generally don't know enough about the real world to produce valid materials.
15. Materials in the process of "becoming" are usually more related to "up front" research and development people than those that have "already become"—yet dissemination strategies almost always relate to "finished products."

**The National Storehouse of Educational Materials/Successes:
Contributors**

Theodore Andrews	Mary Johnson	Allen Schmieder
Robert Ardike	Bruce Joyce	Joseph Schneider
Michael Berger	Susan Klein	William Smith
Joel Burdin	Jori Mark	Francis Sobol
Patricia Cabrera	Karl Massanari	James Steffensen
Paul Collins	Maira Mathieson	William Tikunoff
Ned Flanders	Donald Medley	Wilford Weber
Daniel Ganellis	Fred McDonald	Margaret Wiesender
Carl Grant	Donald Orlosky	Philip Winne
John Hansen	Lorraine Poliakoff	Sam Yarger
Donald Hawkins	Preston Royster	

BEST COPY AVAILABLE

PART V

SOCIAL CONTEXT

17 The Politics of Teacher Competence

Harvey B. Scribner
Leonard B. Stevens

Since the matter of politics generally is not raised in most discussions of teacher competence, the title of this paper perhaps requires a brief explanation. "Politics" as used here does not refer to the partisan activities of political parties or the influence that political officeholders may have occasionally or regularly in issues of education. Rather, "politics" refers to the activities—which, though nonpartisan, are keyed to identifiable interests and goals and thus are, in the broad sense of the word, political activities—and pursuits of groups and individuals within the arena of education.

American public education is well known for its unpolitical character. Clearly, neither the public schools nor the public universities developed in a political or social vacuum. It is equally clear that there has been a notable and commendable absence of overt political, partisan control of the nation's systems of public education.

From this, however, one should not assume that educational institutions function without a political atmosphere. They do have and always have had a political atmosphere. Local school boards, for example, typically pride themselves on their independence from city hall. Yet all but a handful of school boards are dependent on political officeholders for raising and appropriating public tax funds for schools. Moreover, every school board to some extent participates in the deployment of fiscal and human resources for reasons that more than occasionally stretch into political or quasipolitical considerations. Similarly, while state departments of education typically work hard to develop an insulation from statehouse influence, rare indeed is the chief school officer who does not know his way around the halls where statutes are written and appropriations are made. Nor can those who man-

age the public universities be accused of total innocence of the art of politics. Despite the academy's characteristic disdain for politics, university heads with budgets and construction programs in mind find it necessary to do periodic fence-mending with elected representatives of the public who control the strings to the public purse. Inside the academy more than a small amount of energy is expended on political maneuvering.

In brief, the public education enterprise, despite its unpolitical reputation, is quite political when "politics" is defined in broad terms. Schools and universities affect, and are affected by, external political forces. In addition, education systems, because they are large organizations with important social functions and immense resources, possess a set of internal political dynamics which at times become quite intense. Teachers and professors organize for special ends, and they are not disinclined to display their collective power to achieve their ends. Administrators are no different, except that they have the advantage of power which comes from their position. Likewise, school boards and university trustees have political power, and they use it; in addition, they band together in organizations on the regional and national levels, and these organizations themselves wield a certain amount of influence over the nation's education systems. Even researchers, who often are regarded as the most objective and least political of all inhabitants of the education forest, clearly have their own values and distinctive perceptions of what is and what ought to be. And to the extent that researchers channel and shape their research and their conclusions in keeping with their value system, they, too, participate in education politics. "Man is by nature a political animal," Aristotle told us long before politics acquired its undeserved, distasteful connotation. And while he perhaps did not have educators in mind, the observation clearly did not exclude them, as contemporary history of schools and universities has made quite evident.

There was a report in the *New York Times*' recently that two public agencies in New York City, the Board of Education and the Board of Higher Education, had met together for the purpose of establishing closer coordination of their respective activities and programs. Now, since both agencies have responsibilities in public education and since both spend public dollars, the effort to coordinate surely is commendable, and the *New York Times* was quick to do precisely this a day or so later on its editorial page. As a kind of footnote to the story, it was reported that once the heads of the two agencies had met, the next thing they did was to meet privately with the heads of unionized teachers and faculty members.

The *New York Times*, it would seem, should have commended the agencies not only for their positive instinct to coordinate, but also for their keen sense of pragmatism.

This is not necessarily to condemn this "touching base" syndrome, though one is left to wonder precisely what private understandings were reached by the agency heads and the staff unions and to wonder as well whether any understandings were reached that might impair the coordination effort in some way. In any event, this incident is a clear example of politics at work. The point is worth making that political activity should be recognized as such by those who participate in it as well as by those who observe it.

It is neither statesmanship, as some seem to view it; nor is it an indispensable element in modern public administration, as others sometimes contend.

The issue of teacher competence, like other policy questions in education, has political overtones. Anyone who reviews the professional literature on competence and performance in teaching is struck first by the volume of writing on the topic in recent months. The *Education Index* for one recent 12-month period² makes reference to more than 230 articles on the education of teachers—no small number of them pertaining directly or indirectly to the question of performance or competency. The second striking characteristic of the literature is the influence peddling that currently is going on.

Perhaps "influence peddling" is too strong a way of putting it. Perhaps one should say that the literature abounds with the work of individuals contributing to a controversial debate. Whatever one wishes to call it, there is no lack of information on where the major constituencies stand in the debate *vis à vis* the political questions of a) who is to control the education and licensing of teachers and b) how is competence to be defined. Lines continue to be drawn that make clear who is the behaviorist and who is the humanist; who is the teacher educator and who is the teacher employer; and who is the professional spokesman and who is the public advocate.

One might conclude that education needs a Dr. Kissinger to work out a settlement and achieve a truce—but perhaps not. Perhaps what is needed, before the issue can be settled with any degree of finality, is a set of guidelines for working toward a settlement. Given the present state of the public schools and the educational concerns of the society to which the schools belong, it would seem that those with a role in the design of policies and programs for teacher competency should keep at least the following general points in mind.

First one should be aware of special interests, not disguise or ignore them. The policy designer need not necessarily make moral judgments as to "good guys" and "bad guys," but he ought to be aware of the politics of policy. Since policy affects people, people will exert all the influence at their individual and collective command to shape and control policy in its formative stages so that emerging policy will be in their best interest. Needless to say, the goals of a special interest do not necessarily make for good public policy or for better schools. Often, indeed, the goals of special interests ignore the public interest, though the rhetoric of interest groups might have the public believe otherwise.

In the formulation of performance systems for preparing and licensing teachers, it seems clear that teacher educators, who have long controlled teacher preparation and thus heavily influenced teacher certification, tend to favor quite strongly the retaining of their favored position. This does not mean they are against systems that ostensibly will enhance teaching performance. But it does mean that they probably will have a very special interest in retaining as much control as possible over teacher preparation programs. The teacher educator may typically be biased toward a belief that the center of gravity of teacher education belongs on the university campus, in contrast to the view of many school administrators who favor

school based teacher training

Similarly, the longstanding objective of organized teachers to control the teaching profession, much as attorneys and medical doctors control their professions, is readily apparent in the current debate over teacher performance. Organized teachers, in desiring to professionalize teaching, argue that controlling the gates to the profession is essential to the making of teaching into a true profession. Whether the organized teacher's affiliation is with the National Education Association or the American Federation of Teachers makes little difference; the desire to control the profession transcends the matter of affiliation. One president of the NEA, sometimes viewed as less aggressive than the AFT, said a few years ago that the NEA has long dreamed of: controlling "who enters, who stays in, and who leaves the profession." Further, he went on, "Once this is done, we can also control the teacher training institutions."³ It seems unlikely that this expressed desire would be rejected by teachers who choose to affiliate with the AFL-CIO.

The policy designer clearly should be informed. He should be cognizant of the varying views on teacher competence, and he should not write off views on the ground that they emanate from a special interest organization. But he should consider the source—which is to say that he owes it to the public to consider the motivation beneath the position. The views of organized teachers, teacher educators, or school boards obviously need to be considered. But so, too, should hidden political agendas be identified, and the policy designer should be surprised when he finds no ulterior motivation at work.

Second, one should be aware that public education is beyond the era of seemingly unlimited expansion and is well into a period, indefinite in length, of flattened curves. A teacher shortage has turned into an overabundance of teachers. Public dollars for education, whatever the governmental source, are diminishing. Student enrollments are likewise flattening out and decreasing in some instances.

Clearly, there is meaning in this development for those interested in teacher competency systems. School boards are interested more than ever in teacher productivity. They are turning, also, for the first time in recent memory, to the disemployment of teachers. In more than scattered cases, teachers are being laid off much like airline pilots and automobile factory workers for lack of funds, fewer students, or a combination of the two. The policy question therefore arises—if performance or competence is to be the standard for employment, should it not also be the standard for disemployment? For if not, are not schools open to the charge of dual standards—performance as the standard for employment, and "last hired—first fired" as the standard for dismissal? Needless to say, there will be sharp disagreement as to the proper solution. Organized teachers will probably argue hard that where some teachers are to be dismissed, those with longest service deserve the most protection. It should be equally needless to note that seniority has little, if anything, to do with performance, and thus this argument essentially is an antiperformance stance. There is, of course, another side to the coin. A school board intent on saving dollars and with no concern for the rights of teachers will quickly conclude that the greatest savings

are to be made when the most senior teachers—those on highest salaries—are dismissed, as opposed to younger teachers on lower salaries. This stance, too, is antiperformance; therefore, the question remains concerning how to shape performance systems that apply both to the entry of teachers into employment and to the involuntary exit of teachers from employment.

Third, in the design of competency systems, one should keep in mind the growing mood of the public for a stronger hand in the governance of schools. More than a few public school parents are correctly acting like consumers. They are asking for accountability systems. They are demanding effectiveness in educational programs. They want choices for their children. In New York City and Detroit, and in other cities as well, parents are deeply interested in the decentralization of school power. In the case of individual schools, parents are demanding closer involvement in decisions that affect school management and the spending of tax dollars. However much some spokesmen for organized teachers may dislike the idea of parents coming into schools to view the performance of teachers⁴ and however difficult parent involvement may make life for school principals and school superintendents, the fact remains that parents are getting justifiably tougher in their demand to be involved and increasingly savvy in pursuing this objective.

It seems that performance systems for preparing and licensing teachers must take account of the likelihood that parents likely will move increasingly into school decision making. In New York City, for example, parents elect community school boards, they have a specific role in the employment of school principals, they have the right by policy to be consulted in the hiring of community school superintendents, and they have one small foot in the door in tenure decisions.⁵ Most important, perhaps, they have an established channel for appealing grievances against local decisions to the highest level of the city school system, and the short history of the grievance process has shown that parents have scored more than a few points by taking their case to higher authority.

Given such precedents, it would seem reasonable to speculate that parents will want and will insist upon a growing role in the assessment of teacher performance. Yet in the current debate over competency systems, while there is frequent concern for the desirability of participatory planning by organized professionals, school boards, state authorities, and teacher educators, there is little, if any, consistent concern for also including organized parents in the planning. Parents, admittedly, are less cohesively organized and thus less easy to identify as representative than the other major parties to the debate. But ignoring the valid interest of parents in the systems that ultimately will emerge is, perhaps, to court a serious problem. For if parents are not involved in the planning of teacher competency systems, especially those that deal with the matter of licensing, then the possibility is more than remote that parents will tend to view the new systems with more than passing skepticism. In brief, would it not be both ironic and tragic if policies are developed for the purpose of promoting higher competency in teaching through the systematic involvement of all interested parties except parents, whose children are to be the direct beneficiaries of better teaching?

Fourth, teacher competency systems should be designed in the knowledge that the American school—however resistant to reform it is and however similar it is regardless of geographical location—is nonetheless an institution that is changing and that is characterized by local diversity.

If performance systems are to be levers for school reform—and they can and should be—they should be designed on the assumptions that school reform will accelerate over time and that teacher performance is subject to varying definitions in keeping with local priorities and needs. There are such things as alternative schools, and their number is on the increase.⁶ There is the distinct possibility that public education will come to adopt the notion of "public schools of choice."⁷ Responsible efforts are underway to expand external diploma opportunities for high school students and external degree programs for college students. Expansion of various forms of learning-by-doing and learning-through-experience—whether in the form of unconventional apprenticeships or more conventional work-study programs—seems likely in the coming years. The very definition of "teacher" is in a process of evolution as more and more adults who are not professionally trained, full-time educators enter the classroom in the form of parent volunteers and occasional instructors of specific subject matter, from home buying to creative writing. All these trends are altering, slowly but inexorably, the meaning of education, the boundary of school, and the definition and role of teacher. These trends should be of critical interest to those who undertake the highly complex task of stating what teacher competence is, how it is to be systematically created, and the measures of competence. For it would seem self-evident that as schools are changed, so, too, will the competencies needed for effective teaching be altered.

School reform moves with halting steps. Schools—whether in Utah or New York, Minnesota or Texas—tend to look very much alike. But status quo and sameness are not the whole story. If the movement for teacher competency takes as given the assumptions that schools will remain in a relatively unchanged form over time and that teacher competency will be defined in relatively the same way despite locality, then the movement, it would seem, is based on two very dubious, uninspiring assumptions.

Fifth, and last, one should keep in mind the central purpose of competency systems, which is to improve the educational lot of the young. The point is obvious, and all will readily agree with rhetorical, politically benign statements that "competency based teacher education is not an end in itself" and that the ultimate objective is "the improved delivery of educational services."⁸

Delivering on the promise, however, will be quite another matter, and one which will require policies with teeth. The clear implication of competency systems is that competency at some point, in some fashion, will be assessed; and not only will performance be assessed, but where it is found wanting, something will be denied to someone in the name of public protection from sub-par performance. The task for the policy designer is a difficult one. In one sense, it is to ensure that the policy which emerges is not so acceptable to all interested parties—teachers and teacher educators in particular—that "the improved delivery of educational services" is less than a certain result.

What, then, should be the guidelines for developing teacher competency programs and policies?

The first, it would seem, should be to strive for flexibility. Teacher competency systems should allow for multiple definitions of competence and multiple methods of assessment. Clearly, teacher competence has one set of concrete meanings for Chicano families who want for their children a formal education that is bicultural and bilingual. Competence may have other sets of meanings for black parents in central Harlem, for the middle class families of Shaker Heights, or for the blue-collar families in the mining areas of Pennsylvania and Kentucky. This is not to speak indirectly in favor of dual standards of expectation for children, nor to encourage the continuation of ghetto schools, nor to deny that some competencies will have universal application. It is merely to say that different localities will demand some different competencies from their teachers—and that competency systems should not only tolerate this diversity but encourage it.

A second guideline should be an inherent capacity for self-reform. Unless competency systems are to be a new orthodoxy as impervious to change as the course-accounting system of teacher education which Dr. Conant found to be bankrupt more than 10 years ago,⁹ yet which still is with us, mechanisms for review and renewal should be built into the new performance based systems.

A third guideline relates to the earlier discussion of school reform. It is hoped that the designers of competency systems will make the value judgment that learning opportunities are enhanced dramatically to the extent that formal school programs recruit and use artists, writers, craftsmen, and other skilled adults in the teaching of the young on a regular, systematic basis. The issue, in brief, is whether competency systems can and will encourage a broader definition of who may teach, or whether they will encourage schools to remain inflexible in the use of human resources.

A final guideline deals with the matter of control. Teacher education, it has been said, "is in a unique bind, caught as it is between the forces of public control, academic control, and control by the public schools."¹⁰ While this is true, it should be added that the bind has been characterized by rather strict separations—the universities controlling the preservice education of teachers, the state agencies administering the licensing apparatus on the basis of academic credits and attained experience, and the public schools operating inservice programs often consisting more of shadow than substance. The proposed competency systems, with shared planning and perhaps shared control, too, include the prospect, as some observers have noted, of closer ties among teacher educators, school systems, and organized teachers. Few, it has been said, "will argue with the desirability of this goal."¹¹

If what is meant by this is that teacher education, thanks to competency systems, will be more rational and more connected to the nation's public school classrooms, then few should argue. But perhaps there is room for a question if the result of these new ties is to be the inevitable thrusting of near-complete control of teacher education into the hands of professional educators, whether at the university level or the level of the classroom, or a combination thereof. One of the fundamental problems with public educa-

tion, one tends to conclude, is its overdependence on professionals. One recalls the observation of Dr. Conant, who certainly was not a member of the New Left, that "what goes on in schools and colleges is far too important to be left entirely to the educators."¹

Clearly, there is need in public education for professionals with higher levels of performance and higher professional status. Exclusive control of the teaching profession, however, is more than a professional matter, it is also a matter of public policy. The case often is made by teachers that they are professionals as much as doctors and lawyers, and therefore they should have comparable professional power to regulate and control their own ranks. This argument, however, selectively ignores two central factors: a) teachers are public employers and b) their clients are in the classroom by law and not by voluntary choice. These are more than incidental points; indeed, they constitute powerful arguments for greater public control of teacher education and for greater parent participation in the planning and operation of teacher competency systems. To raise this point is not to say that teachers have no legitimate interest in shaping teacher education and teacher competency systems. It is to say that the public has an even larger, more fundamental interest in these matters because of the obvious facts that the schools are public agencies and the young are compulsory students. Yet, obvious as they are, these points tend to get lost in the intricate debate as to what precisely it is that constitutes competence in teaching and how exactly this competence can and should be measured.

Indeed, the answers to these questions—what constitutes competence, and how is competence to be measured—will reflect the value system of the individual or the group to whom the question is put. This, in essence, is the politics of teacher competence. And this is the reason that the parent deserves as much or more of a role in the design of programs and policies for teacher competence as the professional.

FOOTNOTES

- ¹ Education Boards Vow Cooperation. *New York Times*, February 24, 1974, p. 30.
- ² *Education Index* (July 1972-June, 1973).
- ³ George D. Fisher, in *Addresses and Proceedings of the 108th Annual Meeting*, NEA, Vol. 108, San Francisco, June 30-July 6, 1970, p. 12.
- ⁴ Albert Shanker, president of the United Federation of Teachers, is among those who have expressed opposition to the notion of parent evaluation of classroom teachers. See UFT's advertisement, *Where We Stand*, *New York Times* (March 26, 1972). School Decentralization Called Disruptive at UFT Parley Here. *New York Times* (March 10, 1974), p. 47.
- ⁵ *Parent Associations and the Schools* (New York: Board of Education, 1971).
- ⁶ Leonard B. Stevens, *Alternative Education with the Public Schools*, Croft Leadership Action Folio 59 (New London, Conn.: Croft Educational Services, 1973).
- ⁷ Mario Fantini, *Public Schools of Choice* (New York: Simon & Schuster, 1973).
- ⁸ Benjamin Rosner and Patricia M. Kay, Will the Promise of C/PBTE Be Fulfilled? *Phi Delta Kappan*, LV, 5 (January, 1974), 290-95.
- ⁹ James Bryant Conant, *The Education of American Teachers* (New York: McGraw-Hill, 1963).
- ¹⁰ Harry Bowes, *AACTE Yearbook 72*, Vol. 1 (1972), 59.

BEST COPY AVAILABLE

¹¹ Rosner and Kay. "Will the Promise of C/PBTE Be Fulfilled?" pp. 290-95

¹² Conant, *The Education of American Teachers*

18 The Decision-Making Process in Planning and Implementing Education Programs that Meet the Needs of Our Diverse Society

Carl A. Grant

This paper on decision making consists of three general sections: a rationale for educational decision making, a clinical model of action, and a report from a national survey. It is written to provide the reader with a clearer understanding of the issues and problems of planning and implementing a CBTE program that reflects and responds to the multicultural population of our society.

Rationale

The United States is a nation whose citizens have constantly desired and demanded the right to full participation in the decision-making processes which influence their way of life. In 1776, the original thirteen colonies rebelled because they were forced to submit to taxation by the mother country, while being denied any meaningful representation in those governmental bodies which had the power to decide both the extent of the taxes to be levied and the uses to which the collected taxes would be put. In 1955, in Montgomery, Alabama, Rosa Parks ignited a social and economic revolution by not acquiescing to those laws and regulations--legislated and enforced by others--that denied her the right to participate fully in American society.

There is similarity between these two historical elements. Both events call attention to the desire and demand of American citizens to fully par-

ticipate in the decisions that influence and have impact on their way of life. Both events are also examples of the strategies used by Americans to alter the decision-making processes which deny individuals equal representation and, therefore, equal participation in their government.

The history of education cannot be examined separately from other historical events, since education is an integral part of the American culture. Today the educational system is experiencing challenges comparable to those issued in 1776 and 1955. People representing various role groups—students, teachers' associations, and the economically and socially disenfranchised—are demanding their right to fully participate in making those decisions which govern or influence their lives. Full participation implies parity—the equal partnership of diverse role groups and cultural groups sharing in the making of decisions which affect their education and their social and personal growth.

The challenges such a partnership poses for the educational community are formidable and long overdue but should prove to be a rewarding endeavor. The necessity for such a partnership in education decision making is addressed in the final report of the Higher Education Task Force on Improvement and Reform in American Education (HETFIRE). The report states that

We, the Task Force, believe that partnership in teacher education, manifested in different ways in various activities, is the key element in educational reform through teacher education. Individuals are paramount, and what education does for and with larger society will be a result of decisions made by individuals interacting. If one accepts this thesis—that social action is a product of individuals interacting in groups—the degree to which education responds to social needs will depend largely on the extent to which individuals, working together cooperatively, are motivated toward a common goal.

The report further adds:

... that educational policy at whatever level must be based on the needs of the related community. Local assessment of educational needs should determine local educational policy just as the assessment of the educational needs of the nation should determine Federal education policy.

The requirement that schools become more responsive to local needs indicates that it must be the local needs that direct what the schools should do—that policy governing the decisions made and implemented in a school is made at the local level. This requires that policy made at levels increasingly more remote from learner and school (at the district, state, and Federal levels) must be the product of, rather than the prescription for, local educational policy. This notion requires that policy-making "begins" at the local level.

with policy made by more general publics being made to facilitate rather than to constrain the development and implementation of local policy. Good policy-making that leading to reform in the interest of learners - has the following prerequisites

- 1 Those to be involved must have organized a policy making body representing all of the constituencies to be affected. The Task Force has named this the Policy Board
- 2 What should be done must be determined by all to be involved
- 3 Decisions regarding criteria for determining the effectiveness of management and operation must involve all parties
- 4 Commitments to shared responsibility in ongoing evaluation must have been established
- 5 Agreements on qualification criteria must have been established for selecting and retaining those who will implement policy.

HETFIRE is only one of many educational groups to recognize and attempt to respond to the present educational needs of equal representation and equal participation in education decision making. Within the last 2 years, there have been a significant number of publications in journals and a significant number of conferences which have treated educational reform, governance, and collaboration. By and large, this treatment has been more exhortation than action. Power is not given; it must be taken. It cannot be assumed that because we hold conferences and write articles about governance, parity, and collaboration in regard to educational decision making that these concepts will be implemented. A case in point is a phrase in the Declaration of Independence, nationally adopted almost two hundred years ago. The following words have still not come true for all Americans:

We hold these truths to be self-evident—that all men are created equal, that they are endowed by their Creator with certain inalienable rights; that among these are life, liberty, and the pursuit of happiness.

In any honest, sincere discussion of the decision-making process in education, attention must also be given to another point—those who participate in the decision making process must represent the racial and cultural diversity of this country. It cannot be taken for granted that just because various role groups are represented, that the racially and culturally diverse populations of this country are also represented—for most often they are not.

Most of the types of role groups which have been historically identified as necessary components of any educational decision making body—college administrators, Federal and state officials, leaders in teacher organizations—are those who have been most inaccessible to nonwhite, non-Anglo-Saxon people. In addition, decision making bodies in education have not included all types of role groups. Community representatives constitute one such group that has largely been ignored.

Clinical Model of Action

What follows is an outline of a workshop that can be utilized as a method of helping educators to better understand the actual process of planning and implementing CBTE programs that respect multicultural society. The four-part model is based on the premise that actual participation in the process of decision making is as necessary as theory and good intentions.

Part I: The Setting

After being given the rationale for the clinic (see preceding section), the participants will be given a scenario of a college undergoing change. The Scenario is designed to provide the participants with one possible example of a college attempting to embark on a new multicultural CBTE program. The scenario also provides clinic participants with a number of character roles that can be used for actual practice in decision making.

SCENARIO

State College is a multipurpose institution in terms of the diversity of its program of studies and the diversity of its students. It is located in an urban, cosmopolitan city with a heavily diverse population of Anglo-Saxons, Blacks, Orientals, and Puerto-Ricans. Since the Brown decision in 1954, State College has dedicated itself to meeting the needs of the community by making its educational programs available to all people regardless of race, color, creed, sex, or national origin. It is a 4-year degree-granting institution with master's degrees being offered in Elementary Education, Early Childhood Education, and Social Work. The total educational program is organized into 10 divisions, one of which includes the Division of Teacher Education.

Teacher Education. This division trains approximately one-third of the total number of graduates of the college. The students are provided with theory and practice necessary to help them become "proficient" elementary or secondary school teachers.

The Division of Education includes six departments:

Early Childhood Education

Elementary Education

Secondary Education

*Health and Physical Education**Audiovisual and Educational Television**Educational Disabilities*

Each department is headed by a department chairperson, and the total division is headed by a division chairperson who is a recent addition to the faculty, recommended highly by the dean of the college's academic affairs for his/her competence in affecting change.

Instructional Program. Though the graduates of the Division of Education have been certified and highly recommended in some area schools, there is general dissatisfaction with the teacher education program and its relevance to the social, economic, educational, and cultural aspects of the community. The educational program is noted for its competent instruction in four or five curricular areas. The students are exposed to proficient lecturers and adequate multimedia. The program is designed around separate subject areas, usually characterized by a series of three credit-hour courses with autonomous instructors who focus on their own competence and interest. These curriculum areas are based on independent goals and objectives. There are innovative practices such as microteaching, interaction analysis, a clinical approach to supervision, independent study opportunities, an inner-city tutoring program, and a bilingual tutoring program. The majority of the academic program is campus centered.

Though these innovative practices and traditional elements are yielding positive results, they never fit together in a way which gives maximum payoff to the students or the community. Thus, the division chairperson and the Elementary Education Department chairperson decided to look for ways to put the pieces together into a program design based on careful planning, the immediate and relevant needs of culturally diverse student teachers, and a multicultural school population. They decided that some of their basic assumptions about teacher education, previously based on tradition, needed rethinking in relation to teachers' roles and to what and how we want children to be and to become. After observing several teacher education programs and many models of CBTE, they decided that one way was to begin looking at teacher behaviors, attitudes, and understandings that facilitate pupil growth physically, emotionally, and culturally.

It is evident that State College is involved in a stage of educational reform. The success of this reform effort will depend upon the inclusion of diverse role groups and cultural groups; it will also depend on the interrelationship of those elements essential for planning and implementing an educational program that meets the needs of a multicultural society.

Part II: Planning the Program

After the clinic participants have had an opportunity to read the Scenario, they will be instructed to select volunteers to come before the group and to portray roles either identified in the Scenario or suggested by it, which

would be necessary and/or helpful in the initial planning of the program at State College. The clinic coordinator is trying to determine whether the role players will choose to plan their program changes using only the characters associated with the State College or whether they will choose to reach out and involve other community members (i.e., lay individuals and/or groups, school district personnel, labor groups, or possibly state and Federal officials). The coordinator will also be observing whether or not the participants address any of the following questions:

1. Who constituted the program staff established to study the feasibility of developing the program? Were minority group members on this staff?
2. Should an educational needs assessment be made? What educational needs should be considered in planning the program? Was an assessment of minority educational needs made?
3. What ideas or suggestions were solicited concerning multicultural education? Which individuals or groups were asked about this?
4. Should the program seek to change or influence institutional goals and/or objectives? Should goals and objectives relate to the problems of cultural diversity?

The roles that the participants select and the issues they address in the initial planning for the program will be listed. The coordinator will not provide feedback at this time

Part III: Implementing the Program

The participants are now ready to start implementing a multicultural CBTE program as suggested by the Scenario and as initially planned (see Part I). The role playing will be continued and the participants will be allowed to add or delete roles depending on what they learned from the planning stage or from what they expect they will need for implementing the program

Again, what is primarily being looked for here by the coordinator is the choice of roles and the issues they address concerning the implementation of the program. Some of the questions the players might raise are the following:

1. What criteria should be used for admission and retention of students in the program? Is there provision for recruitment of minority students?
2. Are teacher competencies specified which relate to understanding cultural diversity? Are attitudinal changes and modifications of

teacher behavior included in the CBTE program? Should minority group members be involved in determining what attitudinal changes are needed?

3. What roles should students and faculty have in decision making? Are there provisions for staff development in relating CBTE to needs of multicultural education?
4. Who should select and design instructional materials? Are instructional materials identified according to some criteria relative to multicultural education?
5. What provisions should be made for continuous feedback of information from all participants in the new program to the assessing agent and/or program staff? What plans can be made for program modifications based on feedback? What provisions can be made for feedback and change from minority group members?

The roles adopted by the participants and the issues they address in implementing the program are again listed. No feedback is provided at this time.

Part IV: Small Group Discussion

Since only those who have participated in the role playing will have had an active part in the clinic up to this point, all those attending the clinic will now be invited to join the proceedings. This will be accomplished by having at least one of the role players join a group of not more than six other individuals. Each of these minigroups will then be given an adequate amount of time to further refine the issues that were raised by the role players regarding the planning and implementation of the multicultural CBTE program. A brief oral report of these findings will then be made by a spokesperson from each of the groups to the entire gathering.

Part V: Feedback and Evaluation

The roles adopted and the issues raised by the participants in Part II and Part III, as well as the additional input provided by the minigroups in Part IV, will be summarized by the clinic coordinator. The coordinator will compare and evaluate the results of the clinic with the findings of a recently completed national survey of CBTE programs.

Report on Survey Results

Very little research has been done to assess the effectiveness of CBTE

programs or to determine the characteristics of design, planning, and implementation which result in that effectiveness. This paucity of research has meant that those involved with the development and conducting of the programs are forced to operate in the dark, with few empirically tested guidelines to assure that their programs will succeed in attaining the desired ends. To remedy this situation, it is imperative that measures of various program characteristics be developed and examined in the light of their effect upon educational outcomes.

One of the few studies so far conducted toward this end is a survey conducted by Grant and Calhoun (1974), which investigated conditions in 87 colleges across America which are in the process of setting up or conducting CBTE programs. One major purpose of this survey was to examine characteristics of these colleges in relation to their effect upon the planning and implementation of the programs and their approaches to multicultural education.

Toward this goal, the Grant and Calhoun study employed rigorously constructed scales to measure the amount of input into the programs from students, minority groups, and the community at large, the pluralistic content of the curriculum, responsiveness to minority needs, the amount of attention devoted to careful planning, and the degree to which administrators perceived their programs as instruments of educational change. One of the more striking findings of this research emerged from a factor analysis of the intercorrelations among these instruments—all of which loaded rather heavily on a single factor. This result indicates a strong degree of cohesiveness among the program characteristics, wherein a program which is strong in any one of the assessed areas will probably be strong in all of the other areas as well. The factor underlying this cohesiveness was tentatively interpreted as quality of program.

When the components of this factor were examined in the context of various characteristics of the institutions supporting the program, such as size, selectivity of admissions policies, library size, the ratio of Ph.D.'s on the faculty, etc., it was discovered that there is virtually no relationship between the success of the program and the quality of the supporting institution. This finding seems to imply that careful attention to the interrelated variables which determine the quality of a CBTE program can result in the development of a good program at any college, almost regardless of the overall finances and facilities of the college.

Within the intercorrelations among the scales, it is noteworthy that one particularly strong relationship was found. The extent to which minority group members have input into the program is correlated ($r = .76$) with the adequacy of the program in meeting minority needs. This relationship is of such magnitude as to suggest that institutions whose programs do not receive input from minority group members will not provide a program adequate to the needs of these people.

In addition to these findings, tabulations of responses to several other items in the questionnaire provided results of interest to the present discussion.

- 1 A large number of the colleges and universities surveyed did not

establish any criteria for determining who would participate in the initial planning

- 2 Over a third of the institutions reported no community involvement in planning most aspects of the CBTE program. Those community representatives who took part did not include many minority members.
- 3 In a large percentage of the programs, minority group members were not involved in determining proposed attitudinal changes and modifications of teacher behavior.
- 4 Only one-half of the institutions have provisions for the recruitment of minority group students into the CBTE program.
- 5 The majority of the available instructional materials were not identified according to criteria relative to multicultural education.

It is clear, however, that the Grant and Calhoun study represents only an exploratory venture into the sea of open questions which surrounds CBTE. Extensive further investigation is an absolute necessity if educators are ever to conduct teacher education with any confidence that they will arrive at the goals they set out to reach.

Summary

Planning and implementing an educational program that reflects and responds to the multicultural population of our society would be an exciting, challenging endeavor. Indeed, if this idea could be realized, we would have taken an affirmative step toward rectifying a system which now denies many of our citizens their fundamental right to representation in any decision-making process that seriously affects the quality of their lives. Racism is still serving as the keeper of the gate, consequently most programs sweep concepts such as parity and pluralism under the institutional rug. It will take courageous and dedicated individuals to implement such a program because of the overt and covert racism that permeates every aspect of our society. Nonetheless, this does not mean that such a program should not be attempted and cannot be developed.

Finally, it is interesting to note, and is worth reporting again, that one significant finding of the study is that if educational programs are to respond to the needs of minority group people, the programs must have input from minority group people. If we are serious about having a program that responds to our diverse society, we must not fool ourselves and think we can get by with a token individual or two; we cannot. To reform an educational program to meet the challenges and needs of our diverse society, we must have individuals from different racial and role groups fully participating in the decision-making process. To reject this idea is to further perpetuate racism.

FOOTNOTES

¹ *Final Report of the Higher Education Task Force on Improvement and Reform in American Education*. George W. Denmark, chairman (Washington, D. C.: AACTE, 1974), pp. 4-5

² *Ibid.*, pp. 8-9

BEST COPY AVAILABLE

PART VI

INSTITUTIONAL CHANGE

19 **Some Unscholarly Views on Institutional Change**

Don Davies

The competency based idea has become a movement. It has all the standard characteristics—confusion, proliferation of ideas and activities differing schools of thought, lots of energy and talent, strong advocates, and outspoken critics. I'm pleased that the idea has "taken off." However, there is one central issue about the direction of the movement that should be addressed.

The question is this: Will competency based teacher education remain largely a research and development effort, focused largely on changing teacher education programs in colleges and universities, or will it move into the larger arena: teacher competence, the range of things represented by the slogan "accountability," the measurement of success of schools, the involvement of the community in educational decision-making, and the evaluation of professional personnel?

The first direction is relatively manageable and relatively noncontroversial. The chances of succeeding are relatively good—if this is the direction chosen. But, the narrowness of the objective means that the enterprise will be relatively insignificant. The path of highest potential success is also the path to triviality.

If the second direction is chosen, the competency based movement will suddenly be center stage and mainstream activity surrounded by extraordinarily complex political and controversial problems and issues. This direction is obviously the significant one, the one in which activities could have profound impact, but it is also a direction in which chances of success (or to be more precise, of achieving institutional changes) are very slim indeed.

I cannot presume to tell you what you should do. I do not even feel confident about predicting what will happen. I simply want to call the dilemma to your attention.

What I will try to do now is to mention a few points that seem to me to be relevant to your concerns about competency education and institutional change. What follows is a collage of thoughts, not an orderly and systematic analysis or synthesis and certainly not a theory of change.

(1) The first point is a question: Who cares? Those of us involved in teacher education as a major part of our professional lives have a difficult time accepting the fact that people outside of our relatively small group of career professionals don't care much about teacher education or "educational personnel development," if you prefer a broader term. This field of study and work has a very low priority on most university campuses, in most school systems, in state governments, in Congress and the federal educational agencies, among the large professional organizations, and among the public at large. The topic is considered either boring or not very important. My own experiences in Washington over the past decade or so confirm the truth of this assertion. Decision makers in the agencies and on the Hill were concerned about "the teacher shortage" and were willing to support legislation and appropriations that addressed that problem. When it became clear to nearly everyone that the teacher shortage had become a teacher surplus, we faced a very difficult task winning support for appropriations under the Education Professions Development Act. We often heard, "If there is a teacher surplus, why should the Federal government be spending money for 'teacher training' which increases the surplus?" Our persistent efforts to show that most of the funds were actually for inservice staff development and that the main purposes had to do with the improvement of the quality of education in schools and colleges never really penetrated the barriers of skepticisms about "teacher training." The widely held view that training doesn't make a difference was reinforced by the growing view that neither teachers nor schools make much difference.

What I am saying should be borne in mind as you dream about financial support for research and development in competency based teacher education. It doesn't seem likely that massive amounts of money for this purpose will be made available. Excessive ambitions for support will lead to more frustration. I remember some reports in the past 2 or 3 years that talked about investing more than \$300 million in "teaching complexes," \$80 million or more in "protocol materials," and tens and hundreds of millions in other related research and development activities. That's carrying wishful thinking to the point of absurdity, given the general low rank of educational personnel development in the minds of people who make decisions about budgets.

(2) The second point is that changing individuals through training will not produce institutional change. Individual change is a necessary but not sufficient part of institutional or systemic change. This is a lesson that we learned through the experience of the NDEA and EPDA institute and fellowship programs. Helping individual teachers and administrators gain new knowledge, attitudes, and skills is essential but not very productive if the person with new learning is placed back in an unchanged system. The

system will usually win. It was this lesson that led us to develop a broader concept of training in the Urban/Rural School Development program and the ill-fated Educational Renewal plan.

(3) The third point is related to the second. One of my great hopes for the competency based teacher education idea at the outset (way back when the Bureau of Research took the initiative to fund the elementary models projects) was that the professional aspects of preparation would move largely from the campus to the field. It seemed logical that a program emphasizing competencies and competency development would be primarily field based, and would force the development of new and more effective relationships between training institutions and the schools and communities in which the graduates of those institutions serve.

The training complex idea was one significant model. It envisioned the creation of a neutral field training site to which both the university and the school system would contribute and from which both would draw. The problem was that the training complex idea required important institutional changes on the part of both the university and the school system. Both institutions resisted and very little really happened, except for a few very promising pilot projects. The incentives that were offered were not sufficiently strong to create the necessary changes in institutional relationships and the power allocations.

Whether or not my hope that PBTE would lead to having most teacher preparation take place in the field is beginning to be realized, I'm not sure; I'm not even sure whether that hope is widely shared. I continue to hope that it is.

(4) The fourth point is that the individual school is the most manageable and appropriate unit for institutional change in education. John Goodlad has been making this point for a long time now, and my experiences over the past few years support his view. One of the basic ideas in the Educational Renewal plan was that the locus of planning, change, and reform should be the individual school and its community. If the idea of pluralistic local school problem solving ever really caught on, there would be a vast new market for the kinds of competency based materials and training programs that are being developed.

(5) The fifth point derives from my Washington experience. The question is: How will the results of research and development and evaluation be used? (This is really a general question, not one specifically related to research and development in competency education.) This is a straw-man question, of course, because I believe that decision makers don't often use the results of research, development, evaluation, policy analysis, and other academic endeavors. This is a more polite way of saying that the decision-making system intends to be irrational, influenced primarily by political, personal, and pragmatic forces. Better research or better analysis will not correct the deficiencies of the system any more than training teachers better will change school systems. In my most cynical moods I am willing to say that I cannot identify one major and profound decision made by the Office of Education, by the various planning and budget offices, by the Congress that was honestly based on the results of research or program evaluation. I can cite many decisions that were made on other grounds—

partisan politics, educational politics, personal bias, desire to save money, and on and on. My point is that those concerned about research and development and evaluation in education must also be concerned about encouraging, developing, and working for ways to make governmental and institutional decision-making processes more rational—more responsive to information, ideas, and documented experience.

(6) The next point, which has some relationship to the last, is that we haven't made much progress yet on the problems that Research and Development people in education have in communicating effectively with either policy makers or with practitioners. NIE is having more than its share of difficulty with this problem, just as I did when I was on the firing-line. I can't pretend to know all of the solutions to the problem. I certainly haven't been successful in solving it myself. However, one of the first things to be done is to clean up the language—to communicate in plain and powerful English. A second step towards solution is to recognize that most people in educational research still seem to be bound to a top-down program based on the familiar model—research, diffuse, demonstrate, install, evaluate, revise, and begin the process anew. More than we admit, we tend to believe or hope that there is a passive institution waiting to receive and adopt the results of our Research and Development endeavors. Obviously, this is a faulty model of institutional change.

(7) Another point, which I mention only briefly, is that very high priority ought to be given by the Research and Development community in education to studying the process of implementation of change, with special references to the complex matters of incentives. For example, it is clear that the incentives for initially adopting an innovation and the incentives for implementing it—making it work—are very different.

(8) And now, a final point is the question: Who runs the schools? The question should actually be, Who runs the society? The most important issues we face as a people today have to do with the relationships between government, the institutions of the society, and the people. And it's a question that has a lot to do with institutional change. Too many people see institutional change as a technical and professional task. Too many people see institutional change as the responsibility of the social scientists, the educational research and development people, the innovative bureaucrats in Washington, or the "cutting edge" superintendents. Too many people have come to accept the elitist idea that ordinary people cannot be trusted to make decisions in a complex and technologically sophisticated society because they can't possibly understand the consequences of their decisions. Hence, we need the "best and brightest" to make decisions—always, of course, in the best interests of the people or the "national security." I believe we should fight this elitist self-fulfilling prophecy.

It is my view that our most important task as a people is to bring our institutions closer to the people, to make those institutions more responsive to the diverse needs and interests of the people, and to find ways whereby ordinary people can share decision-making powers with the professionals and experts in various fields. The work that I am doing now is focused almost entirely on this task in relation to citizen participation in decision making in the schools. It is clear to me that school improvement, school change,

school reform, institutional change in education will happen to the extent that we learn how to involve parents and the community directly and significantly with teachers and administrators in a process of planning and problem solving, particularly at the level of the individual school. Without this kind of participatory process, change isn't likely to occur (or if it occurs, it isn't likely to last), and what is even more important, it shouldn't occur.

Despite all the concerns and warnings expressed in these hurried comments, I believe that educational research and development people and the competency education movement have a good deal to contribute to this kind of process of institutional change.

20 Task Force on Institutional Change

**James Dyer
Dell Felder**

This task force consisted of from 12 to 15 members (at different times) and decided at the outset to discuss and develop the research issue as a single group. The chairman, Dell Felder, introduced the game play and suggested that the context be broadened to include common issues that influence change in all sizes and types of institutions.

It appeared that two assumptions were made generally by the group (though not actually verbalized until much later). One was that the institution that was referred to was, in fact, the institution of higher education, and the second, that the change referred to was the implementation of CBTE.

Introduction around the group disclosed a variety of representatives accounting for all critical change interest groups with the exception of the professional teacher organizations. The brainstorming technique was used to bring forth a flood of issues, primary areas of interest being increased costs, professional autonomy, administration involvement, and the general relationship currently existing between the university, public schools, professional teacher organizations, the community, and the state departments. (Note: these will be herein referred to as the groups.)

There were strong feelings about the magnitude of change to be instituted in the initial stages, the commitment of the various authority figures, and the fact that change, as such, could never be "rammed down anyone's throat." Many of the group continued to clamor for "more proof" before change, though others recognized the need for establishing valid research in actual experimental programs before such proof could be forthcoming.

The recommendations for research issues and their group-assigned

priorities are as follows

1. What, specifically, can IHE & other groups learn from each other?
2. Should research efforts be shared by the five groups?
3. What are the prerequisites for collaboration between all five groups?
4. Is a new professional school of education desirable?
5. What are the emerging functions of professional schools of education?
6. Are pioneer (pilot) programs desirable?
7. How do you develop a pioneer (pilot) program to influence the entire system?
8. How is a successful prototype program implemented as *the* program?
9. How do (job) supply and demand influence institutional change?
10. How can threat to faculty and professional autonomy be avoided?
11. What rationale is there for nonchange in all five groups?
12. Is the traditional reward system really a major influence on change?
13. What does institutional change include?
14. Do institutional histories and culture make a difference in developing change?
15. Do educators believe teacher education needs to be changed?
16. Can the Teacher Education Center help develop institutional change?
17. Should change (CBTE) be limited to undergraduate areas?

Feeling was strong that the research personnel should include the interested parties, and that all prior related studies be consolidated and promulgated as soon as possible.

21 Model for Developing a State System for Managing Performance Programs*

Robert C. Burkhart

A

Overview

The results of a study completed by Dr. Robert C. Burkhart, President of Educational Assessment Systems Corporation, and supported by the New York State Education Department indicate that student teaching evaluation forms in use in many New York State institutions of higher education with approved programs in education consist largely of judgments made on the basis of either subjective requirements or subjective criteria. The results also imply that the evaluation of student teachers emphasizes portions of professional education for which formal instruction is seldom provided.

By supporting this study, the Division of Teacher Education and Certification of the New York State Education Department was attempting to gauge the present state of the evaluation of student teachers in approved programs in education. It was felt that publication of the results of the study would allow teacher preparation institutions to measure their own progress toward the implementation of a competency based teacher education program.

Conclusions drawn from the results of the Burkhart study indicated a

Michael King and Mary Van Ryn, both of the New York State Education Department, assisted in preparing the above article for publication in this volume.

significant contrast between a majority of the student teaching evaluation forms now in use and the assumptions and principles outlined in "The Educational Rights of Students," a statement prepared by Burkhart and Mike Van Ryn, Chief of the Bureau of Inservice Education, Division of Teacher Education and Certification, NYSED (see Table 21.1).

To conduct this study, data were gathered from the student teacher evaluation forms used in 89 New York State institutions of higher education with approved programs in education. The data were analyzed to determine the emphasis placed upon various portions of student teaching and the assessment standards employed in making judgments about the performance of the student teachers. A result of the analysis: a "Representative College Student Teacher Evaluation Form" (see Table 21.2), a prototype of the forms now in use, was constructed and analyzed.

The Study

The 89 collected forms were analyzed to determine the total number of discrete evaluations required by all forms. A total of 3,191 discrete items were identified and 3,015 were coded according to the Educational Assessment Systems Corporation Analysis System. 176 items were discarded because they were too general to be classified.

The EASC Analysis System is a description in list form of the functions required of a student teacher (see Table 21.3). For the purposes of this study the instrument was presumed to be exhaustive. The Analysis System consists of three major areas: The Teacher-Learner Pre-Instructional Functions Area, the Teacher-Learner Instructional Functions Area, and the Teacher-Learner Professional Profile Area. The areas are divided into categories as follows: Pre-Instructional Functions Area, four categories; Instructional Functions Area, four categories; and the Professional Profile Area, two categories. The ten categories are further divided into a total of 57 subcategories. Each subcategory has its own distinctive number. The distinct number of the subcategory was the number used to code the discrete evaluation items discovered in the 89 evaluation forms.

The decisions regarding classification were made by individuals who had received an orientation in the use of the Analysis System. Each item was rated separately by more than one judge and the inter-rater reliability factor was found to exceed .90.

To construct the "Representative College Student Teacher Evaluation Form" (REF), it was necessary to complete two independent analyses of 3,015 items. As a group, the items, coded according to the subcategories in the Analysis System, were tabulated to determine their frequency of use in the 89 forms. While number 23 of the subcategories was not mentioned, the remaining 56 subcategories showed a frequency ranging from 1 to 500 times on the 89 forms.

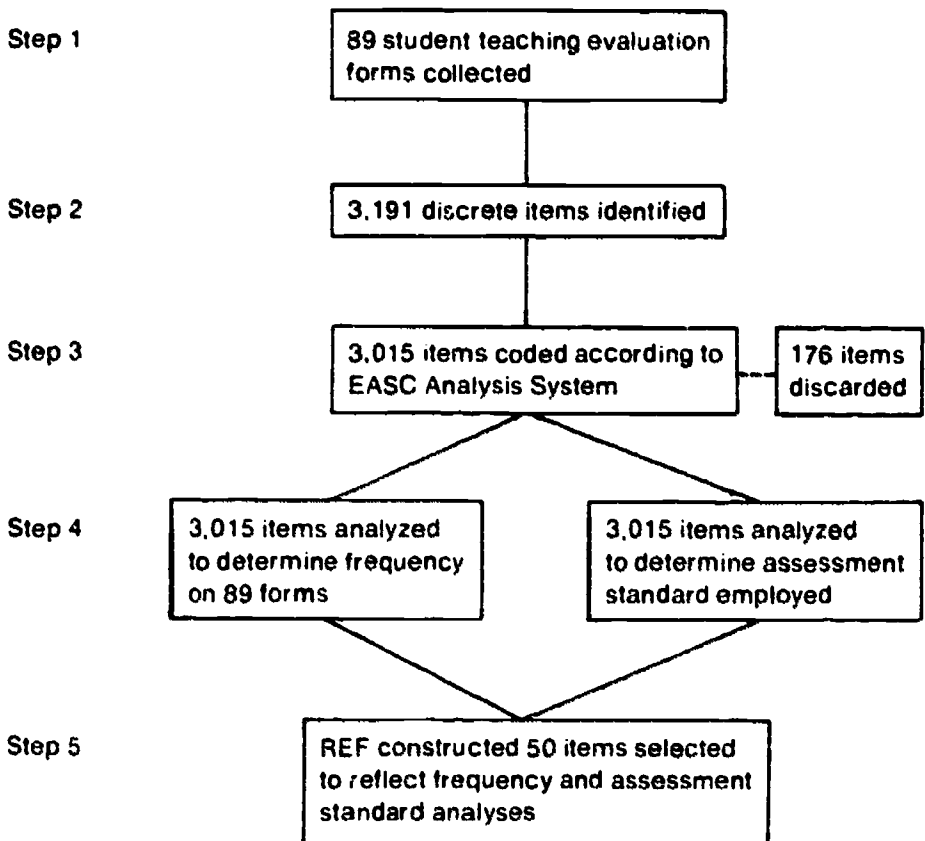
In the second analysis, each of the 3,015 items was examined to determine the assessment standard employed. The following is a list of the standards found to have been used, the code identification, an explanation and example of each, and the percentage of the total number of items repre-

Standard	Code	Explanation/Example	Per Cent
Category	C	The name of an area in which an assessment is to be made: "Questioning."	14%
Definition	D	The name of an area in which an assessment is to be made and mention of the distinguishing characteristics of the area: "Subject matter comprehension: the understanding of concepts through teaching."	1%
Subjective Requirement	SR	A statement that requires a subjective judgment about achievement: "Effective use of black-board."	72%
Subjective Criteria	SC	A statement that requires a judgment about a level of achievement: (scale required) "Acceptance of constructive criticism."	7%
Performance Requirement	PR	A statement that requires rating based on behavioral occurrences: "Accurate reports handed in on time."	6%

Based on the two analyses, 50 actual evaluation items were selected to constitute the REF. All subcategories that had appeared at least once on more than 50 percent of the 89 forms were represented according to their relative frequency. The 50 REF items also reflected the percentages determined by the second analysis. For example, "subjective requirement" was found to have been employed in 72 percent of all evaluations; therefore, 72 percent of the REF items employed a subjective requirement as an assessment standard.

Figure 21.1, below, graphically represents the processes involved in the study, as described above, from the collection of the student teaching evaluation forms to the construction of the REF.

Figure 21.1



The following charts highlight the primary findings of the Burkhardt study. Chart 21.1—Emphasis on Areas of Assessments in REF Determined by Frequency of Item Occurrence—presents information about the emphasis placed on various portions of the REF. The number of evaluations per sub-category and category and the percent of the whole represented by these figures by category and area are presented. Chart 21.2—Emphasis on "Subjective Requirement" and "Subjective Criteria" Assessment Standards in REF Determined by Frequency of Use—presents the number of evaluations in one category that employed either a "subjective requirement" or "subjective criteria." The percent of evaluations in one area that employed either a subjective requirement or subjective criteria has also been calculated.

Chart 21.1

**EMPHASES ON AREAS OF ASSESSMENTS IN REF
DETERMINED BY FREQUENCY OF ITEM OCCURRENCE**

EASC Analysis System Categories	EASC Analysis System Subcategory Code Numbers	Number of Evaluation Items Per Subcategory	Number of Evaluation Items Per Category	Percent of Total Evaluation Items Per Category	Percent of Total Evaluation Items Per Area	EASC Analysis System Areas
Preassessment of Pupil Skill and Techniques by Teacher-Learner	1 2	1	1	2%	10%	Teacher- Learner Pre-Instruc- tional Functions
Planning Instruc- tion Objectives and Strategies	2-1	1	1	2%		
Setting Up Pro- cedural Routines for Instruction	3-1 3-2	2 1	3	6%		
Maintenance of a Positive Learning Environment	5-1 5-2 5-3 5-4	1 2 1 1	5	10%	46%	Teacher- Learner Instructional Functions
Conducting Learning Experiences	6-1 6-3 6-4	4 1 3	8	16%		
Knowledge of Content Area and Procedural Skills Used During In- struction	7-3 7-4	5 4	9	18%		
Assessment Feedback --- Remedial Help	8-2	1	1	2%		
Assessment of Continuing Professional Development	9-3	1	1	2%		
Professional Characteristics and Interests of a Teacher	10-1 10-2 10-3	6 5 10	21	42%	44%	Teacher- Learner Professional Profile
TOTAL	18	50	50	100%	100%	

Chart 21.2
EMPHASIS ON "SUBJECTIVE REQUIREMENT" AND
"SUBJECTIVE CRITERIA" ASSESSMENT STANDARDS
IN REF DETERMINED BY FREQUENCY OF USE

EASC Analysis System Categories	EASC Analysis System Subcategory Code Numbers	Number of Evaluation Items Per Subcategory	Number of Evaluation Items Per Category	Number of Evaluation Items Using Assessment Standards SR and SC Per Category	Percent of Evaluation Items Using Assessment Standards SR and SC Per Category	Percent of Evaluation Items Using Assessment Standards SR and SC Per Area	EASC Analysis System Areas
Preassessment of Pupil Skill and Techniques by Teacher-Learner	1-2	1	1	1	100%	60%	Teacher- Learner Pre-Instruc- tional Functions
Planning Instruc- tion Objectives and Strategies	2-1	1	1	1	100%		
Setting Up Pro- cedural Routines for Instruction	3-1 3-2	2 1	3	1	33%		
Maintenance of a Positive Learning Environment	5-1 5-2 5-3 5-4	1 2 1 1	5	3	60%	60.9%	Teacher- Learner Instructional Functions
Conducting Learning Experiences	6-1 6-3 6-4	4 1 3	8	5	62.5%		
Knowledge of Content Area and Procedural Skills Used During Instruction	7-3 7-4	5 4	9	5	56.5%		
Assessment Feedback - Remedial Help	8-2	1	1	1	100%		
Assessment of Continuing Professional Development	9-3	1	1	1	100%	100%	Teacher- Learner Professional Profile
Professional Characteristics and Interests of a Teacher	10-1 10-2 10-3	6 5 10	21	21	100%		
TOTAL	18	50	50	39			

Although the major conclusions are largely self-evident, the most important results should be mentioned and briefly discussed because of their implications.

Chart 21.1 illustrates two major findings: (1) little emphasis is placed on the pre-instructional functions of most student teachers and (2) considerable emphasis is placed on the professional characteristics of the student teacher. While an almost equal amount of weight is given to the Instructional Functions Area and the Professional Profile Area, this is so only because one of the categories in the Professional Profile Area, Category 10, "Professional Characteristics and Interests of a Teacher," constitutes 42 percent of all the evaluations. That is, while the evaluations for the other areas are distributed throughout the categories constituting the area, the evaluations in the Professional Profile Area are almost wholly located within one category. Furthermore, the number of evaluations in one subcategory, number 10-3, is greater than the number of evaluations in any other single category. It is also of interest to note that there are no evaluations from Category 4, "Occupational Responsibilities."

Chart 21.2 demonstrates that the bulk of evaluations employ either "subjective requirements" or "subjective criteria" as assessment standards. There is only one category, number 3, "Setting up Procedural Routines For Instruction," in which less than 50 percent of the evaluations do not employ subjective assessment standards. All the evaluations, or 100 percent in the Professional Profile Area employ subjective standards.

When the information presented on Charts 21.1 and 21.2 is compared, the conclusions are significant. All evaluations in the Professional Profile Area, which constitute 44 percent of the evaluations, employ subjective assessment standards. The one category that does not use subjective assessment standards for more than one-half of its evaluations constitutes only 6 percent of all the evaluations made. While the conclusions are obvious, their prominence should not belittle their importance. Most student teachers are being evaluated by predominantly subjective assessment standards in all areas. Furthermore, the largest single area of evaluation, which relates to professional characteristics, is wholly evaluated by subjective standards. It is not, therefore, extreme to conclude that most of the student teacher evaluation forms used in New York State institutions of higher education with approved programs in education are not framed to insure the application of consistent criteria, the making of fair judgments, or the accurate reflection of the educational programs of the institution.*

As teacher education practices in New York State become competency based, it will be important to assure that the evaluation of students be consistent with instruction and based upon explicit, publicly stated standards. It is hoped that this summary of the Burkhart study will be helpful to the institutions of higher education in New York State as they assess their efforts in converting to competency based education programs.

*A cursory review of the education courses offered by 46 institutions of higher education in New York State that offer approved programs in education located no courses that were wholly devoted to "Educational Professionalism" and found only infrequent mention of that or related topics in course descriptions

Table 21.1

**THE EDUCATIONAL RIGHTS OF STUDENTS ENROLLED
IN TEACHER EDUCATION PROGRAMS**

The educational rights cited below should be viewed within the context of a democratic society that guarantees all citizens certain human and civil rights. Moreover, they should be considered as congruent with those rights held by faculty and institutions and consistent with the principle of public disclosure.

The educational rights of students rest on the following assumptions:

- 1) that achievement evaluation is only justifiable in areas where instruction has been provided;
- 2) that, when instruction is provided, evaluation is essential;
- 3) that an instructional system, if it is to be responsible, must concern itself with the explicit basis for instruction, provide feedback, and publicly disclose the requirements to be met; and,
- 4) that evaluation should be consistent with instruction and be congruent with the explicit diagnosis of students as indicated by their profiles as learners.

Therefore, in a teacher education program, the student has the right to expect the following:

- 1) an instructional program that will help him to acquire a level of competency which enables him to assume instructional responsibilities in a classroom;
- 2) that the instructional program is a direct outgrowth of identified competencies;
- 3) that the competencies he is to attain are explicitly and publicly stated;
- 4) that continuous feedback about his progress will be provided;
- 5) that there will be sufficient opportunity to make progress while involved in the program;
- 6) that the assessment procedure will have publicly stated conditions of performance and designated levels of mastery;
- 7) that the assessment procedure will include a pre-assessment which will allow him to demonstrate his level of performance prior to his involvement (enrollment) in any component of the instructional program;
- 8) that a record of performance will be continuously available to him in a profile form; and,
- 9) that the standards of achievement to receive recommendations for certification are established based on the preceding criteria.

Table 21.2
REPRESENTATIVE STUDENT TEACHER EVALUATION FORM

ASSESSMENT STANDARDS CODE

C-Category
D-Definition
SR-Subjective Requirement
SC-Subjective Criteria
PR-Performance Requirement

1. PREASSESSMENT OF PUPIL SKILL AND TECHNIQUES BY TEACHER-LEARNER

Individualization of learning strategies according to pupil needs
Items Provision for meeting individual differences (SR)

2. PLANNING INSTRUCTIONAL OBJECTIVES AND STRATEGIES

Lesson design, unit design
Items The student teacher develops lessons with consideration for pupil pace (SC)

Pacing, variety of technique, use of <i>unexpected</i> material	Alters pace of teaching to meet pupil needs	<i>Not aware</i> of student response to lesson

3. SETTING UP PROCEDURAL ROUTINES FOR INSTRUCTION

1. Selecting and organizing procedures, equipment, and facilities
Items Selection, use, and care of room, equipment, and materials (SR)

Regulates physical conditions of the room; lights, head ventilation (PR)

2. Routine duties

Items. Accurate reports handed in on time (PR)

5. MAINTENANCE OF A POSITIVE LEARNING ENVIRONMENT

1. Classroom control: direction and procedure
Items Classroom direction and procedure (C)

2. Disciplining disruptive pupils

Items Adequate classroom control (SR)

Just and effective in handling disruptive students (SR)

3. Positive learning approach

- Items. Pupil praise (C)
 4. Pupil participation in learning activity
 Items: Ability to involve all pupils activity (SR)

6. CONDUCTING LEARNING EXPERIENCES

1. Motivation and presentation
 Items. Motivational devices (C)
 Dramatic motivations (SR)
 Imaginative motivations (SR)
 Introduction and stimulation of activity (SR)
2. Questioning and answering skills
 Items. Questioning (C)
3. Instructional materials, techniques, visual aids, audio-visual skills
 Items. Use of two visual aids in teaching a lesson (PR)
 Effective use of blackboard (SR)
 Sufficient use of audio-visual equipment (SR)

7. KNOWLEDGE OF CONTENT AREA, PROCEDURAL SKILLS USED DURING INSTRUCTION

1. Communication techniques
 Items: Voice and speech (C)
 Student teacher vocabulary (C)
 Speaking distinctly with expression and quality (SR)
 Pleasant, well modulated voice pattern (SR)
 Command of English language, good vocabulary (SR)
2. Utilization, comprehension of principles in content area, basic concepts
 Items. Subject matter comprehension, the understanding of concepts through teaching (D)
 Principles in subject matter (C)
 Knowledge of subject (SR)
 Student teacher knowledge of subject matter to be introduced (SC)

<i>Resourceful imaginative use of pupil background to develop content</i>	<i>Works into teacher guide, selectively adds own materials</i>	<i>Non selective, not usually prepared</i>

8. ASSESSMENT-FEEDBACK-REMEDIAL HELP

- Individual and group evaluation of pupils
 Items: Effective use of evaluative devices (SR)

9 ASSESSMENT OF CONTINUING PROFESSIONAL DEVELOPMENT

Openness to critical comments by others

Items Acceptance of constructive criticism (SC)

Seeks suggestions; always willing to accept suggestions	Takes suggestions to a <i>limited degree</i>	Fails to follow suggestions

10 PROFESSIONAL CHARACTERISTICS, INTERESTS OF A TEACHER

1 Role awareness and ethics

Items Dedication to teaching (SR)

Initiative (SR)

Professional ethics in evidence (SR)

Judgment and tact (SR)

Dependability, prompt (SR)

2 Personal and professional characteristics

Items Poise, self-confidence (SR)

Pleasing appearance (SR)

Intellectual curiosity (SR)

Good health, vitality (SR)

Good posture (SR)

3 Affective capacities: enthusiasm, desire, sympathy

Items Openmindedness (SR)

Healthy pupil-teacher relations (SR)

Emotional stability (SR)

Warmth of personality (SR)

Sensitivity, rapport (SR)

Sense of humor (SR)

Patience, self-control (SR)

Sociability, friendliness (SR)

Spontaneity, enthusiasm (SR)

General quality of relations with children. (SC)

Ability to make and maintain contact with children, empathy; ability to relate in imaginative ways; acceptance of and respect for childrens' work; understanding of their need for challenge; reactions to negative behavior (SC)

Relationships are good in all or most aspects	Relationships are <i>adequate</i> but <i>uneven</i>	Poor relations

Table 21.3

**EDUCATIONAL ASSESSMENT SYSTEMS CORPORATION'S
ANALYSIS SYSTEM**

TEACHER-LEARNER PRE-INSTRUCTIONAL FUNCTIONS AREA

- 1. PREASSESSMENT OF PUPIL SKILL AND TECHNIQUES BY
TEACHER-LEARNER CATEGORY**
 1. Development and use of pupil profiles-achievement ratings
 2. Individualization of learning strategies according to pupil needs
 3. Diagnosing pupil problems and needs prior to instruction
- 2. PLANNING INSTRUCTION OBJECTIVES AND STRATEGIES
CATEGORY**
 1. Lesson design, unit design
 2. Establishing objectives
 3. Establishing evaluative criteria
 4. Establishing requirements
 5. Making and selecting tests
 6. Sequencing instructional activities
- 3. SETTING UP PROCEDURAL ROUTINES FOR INSTRUCTION
CATEGORY**
 1. Selecting and organizing procedures, equipment, and facilities
 2. Routine duties
 3. Observation of instruction
- 4. OCCUPATIONAL RESPONSIBILITIES CATEGORY**
 1. Research and projects development of instructional resources
 2. Scoring tests and grading
 3. Providing resources and services—bulletin boards, etc
 4. Co-operation with colleagues
 5. Fulfills responsibilities
 6. Providing teacher assistance
 7. Using consultant personnel

TEACHER-LEARNER INSTRUCTIONAL FUNCTIONS AREA

- 5. MAINTENANCE OF A POSITIVE LEARNING ENVIRONMENT
CATEGORY**
 1. Classroom control, direction and procedure
 2. Disciplining disruptive students
 3. Positive learning approach
 4. Pupil participation in learning activity
 5. Individual work by pupils
 6. Monitoring of pupil interest
 7. Providing pupil assistance
- 6. CONDUCTING LEARNING EXPERIENCES CATEGORY**
 1. Motivation and presentation

2. Teacher lecture and demonstration techniques
 3. Questioning and answering skills
 4. Instructional materials, visual aides, audio-visual usage skills
 5. Supervising pupil skill practice
 6. Interactive problem solving, critical thinking, and discussions
 7. Relational learning experiences
 8. Summary and conclusion; reinforcement and review
7. **KNOWLEDGE OF CONTENT AREA AND PROCEDURAL SKILLS USED DURING INSTRUCTION CATEGORY**
1. Accuracy of information
 2. Adequacy in employment of procedural skills
 3. Communication techniques
 4. Utilization and comprehension of principles in content area, understanding basic concepts
 5. Provision of application procedures necessary for pupil problem solving
 6. Understanding or applying educational philosophy
8. **ASSESSMENT-FEEDBACK-REMEDIAL HELP CATEGORY**
1. Assessment: pupil product, process, and knowledge for diagnostic purposes after instruction
 2. Individual and group evaluation of pupil work
 3. Feedback and analysis of pupil performance
 4. Remedial help based on analysis of pupil performance
 5. Re-assessment of pupil learning, evaluative instruments, and lesson planning based on feedback and remedial help
- TEACHER-LEARNER PROFESSIONAL PROFILE AREA**
9. **ASSESSMENT OF CONTINUING PROFESSIONAL DEVELOPMENT CATEGORY**
1. Preassessment of performance capacities of teacher learner
 2. Self-analysis and self-evaluation by teacher learner
 3. Openness to critical comments by others
 4. Capacity to identify means of improvement
 5. Pupil learning results: achievement and gain
 6. Pupil reaction to teacher learner
 7. Demonstrating ability to progress as teacher learner
10. **PROFESSIONAL CHARACTERISTICS AND INTERESTS OF A TEACHER CATEGORY**
1. Role awareness and ethics
 2. Personal and professional characteristics
 3. Affective capacities enthusiasm, desire, sympathy
 4. Personal involvement in content area activity
 5. Teacher interests, extra-curricular activity, community activity, etc.

B

A NEW STYLE CERTIFICATION INSTITUTE: DATA BANK TRAINING INSTITUTE*

Introduction

The purpose of this Data Bank is to collect critical kinds of information at a low cost, which will be of maximum value for providing pertinent information to the New York State Education Department, school districts, and colleges about the performance and effect of teacher education programs on pupil learning in the public schools. Two types of information have been designated as of utmost importance. The first is the final evaluation forms used by supervisors to assess teacher learner's performance in the classroom. The second is an observation by trained observers of the range of these teacher learners' education activities during their training program. Using this information, the Data Bank charts an institutional profile of the training program's means of assessing student teachers and of their actual activities as learners in the classroom. This profile is used to review, evaluate, and formulate instructional programs in ways which increase the decision-making responsibilities and accountability of associated administrators, faculty, and students. The purpose is to increase the range of learning opportunities provided to the teachers-in-training, and simultaneously to provide a widening range of learning opportunities for public school pupils. The Data Bank System, through a continuous feedback process, allows the training staff at the college and public school level to identify gap areas in instruction and to concentrate in the field on filling them. Once the gap is identified, instructional resources are developed which will enable the teacher learner to provide pupil instruction in this neglected area. The formulation of instruction that is adequate in meeting Data Bank standards requires two steps.

- (1) That the means of assessment is stated prior to instruction, and is formulated in terms which are observable;
- (2) That instruction is provided which fulfills the requirements of these means of assessment. This provides for the educational rights of the teacher learner as adopted by the Advisory Board of the Regents of the State of New York.

Preassessment of Local Institutional Profile

A distinguishing feature of our program is that there is a preassessment profile made of the participating institutions' educational program. This is followed by inservice training based on special needs of the participating institution, in order to provide for immediate positive actions in areas of

*The Data Bank for the state of New York Division of Teacher Education and Certification was designed by the Educational Assessment Systems Corporation, 1432 Amherst Street, Buffalo, New York, 14216

This system is one of several designs now being considered for adoption by the New York State Education Department

need. When desired, a followup profile is provided to determine areas of gain as a result of the work done. This pattern is the model established by the Data Bank for the certification of programs in the State of New York.

The profile will consist of a 2-day visit to the participating institutions, consisting of:

- (1) Sixteen to twenty 30-minute observations made by the EAS staff of the institutions, paraprofessionals, and student teachers
- (2) A collection of information concerning:
 - a. Student teaching and other teacher learning assessment instruments.
 - b. Task descriptions of teacher learners' jobs, plus role descriptions, method course outlines, all available instructional modules, and evaluative procedures.
 - c. Fifteen-minute interviews with 10 student teachers, four co-operating teachers, four method instructors, four supervisors, and administrators to determine a picture of the institutions' governance operations and advisory board, curriculum procedures, program sequence, and management systems for providing feedback on program effectiveness.

The material collected here is necessary in order to provide highly relevant kinds of training experiences which result in immediate program improvements.

After this preassessment visit and development of the institutions' profile, the necessary resource material is organized as follows for the participating institutions' particular needs.

Assessment of Teacher Learners

Training in assessment techniques for teacher learners will take between 4 and 5 hours. It will consist of a general session on theory concerning the relationship of the "Educational Rights of Students" to the means of assessing their performance. There are two major contents to be mastered. The first concerns the statement of performance criteria for assessment and a related report on major errors made by 89 institutions on their student teacher evaluation forms. The second concerns the necessary range of assessment items if a full training program is to provide an adequate base for pupil and teacher learning. Exercises in both content skills are provided, including a review.

(1) *Theory.* This section will be the introduction of the "Educational Rights of Students." This will be followed by an analysis showing how means of assessment must be related to means of instruction, if training programs are to be effective. Specific requirements essential to relating means of assessment to means of instruction are given so that job responsibilities of the teacher learner are specified and training in them is provided. Without meeting the two sets of requirements, no program can operate effectively.

(2) *Statement of Performance Criteria.* This area will specify mistakes frequently made by educational institutions in the statement of performance

assessments for student teachers. These are the "Thou Shalt Not" exercises. This section ends with the analysis of the participating institutions' own evaluation forms.

(3) *Range of Assessments.* This area includes a report on the gap areas in student teacher training programs of 89 institutions in New York State. Also included will be a statement and definition of the 11 Data Bank categories needed for assessment. The Data Bank categories fall into three major groups: (1) prior to instruction; (2) during instruction; and (3) in addition to instruction. There are 55 total items, 41 of which relate to classroom observation. An exercise will then be given in which participants will match statements of assessment with the Data Bank categories. Another exercise will follow, in which the participants will sort their own institutional items for assessment for student teachers under the Data Bank categories. The results from these exercises will be used to identify gap areas in their own institutional student teacher assessment forms.

(4) *Filling Gap Areas.* In the prior two exercises, the gap areas have been identified and the requirements for stating performance criteria clarified. This meeting deals with the formulation of two necessary forms of assessment: convergent types of assessment for which pass/fail types of requirements can be stated; and divergent types of assessment problems which require the formulation of criteria. Success in formulation of both types of means of assessment depend upon identifying specifically the function to be performed by the learner or the elements of the problem to be solved, which distinguish between varying degrees of achievement. This is a critical skill because it determines the requirements of instruction. The final exercise in this session will involve the participant in stating task descriptions for criteria. Included in this exercise is the identification of related job specifications for this task and the development of proficiency and competency tests for determining if the job has been done effectively by the teacher learner. (80 min.)

(5) *Anonymous Evaluation of the Session by the Participants.*

Introduction to the Observation System

Training in the observation system will take between 4 and 5 hours. It consists of a lecture demonstration and video tape exercises of actual classroom situations to show how the observation system is used. The training will also show how to make an analysis of the observation findings. This exercise informs the observer concerning how the observation system provides a basis for necessary decision making regarding the implementation of pupil and teacher learning. The observation system records 41 discrete kinds of classroom activities and skills. They are grouped into nine sections:

(1) *Preassessment.* Determines participants' prior knowledge of observable teacher tasks. Group analysis of the preassessment test will follow for review.

(2) *Goals and Means of Presentation.* Explains the objectives, requirements, and criteria of the observation training session. This will be the means by which the participants will evaluate the success of the workshop.

(3) *Presentation.* Lecture demonstration by experienced staff identi-

tying the 41 items on the observation form. A videotape example of each item will be used to supplement the definition.

(4) *Practice.* Consists of two different observation exercises, viewed on video tapes from two different classroom and content areas. Participants will use the observation form for the first time, in order to familiarize themselves with the pattern of using the form. These exercises provide for the recognition of the same item in different content areas. (30 min.)

(5) *Assessment as Analysis.* Consists of simulated field training. The participants will be broken down into two groups. Each participant in the group will observe four different video tapes of actual classroom situations using the observation form as they would in the field. The tapes will be run continuously. (80 min.)

(6) *Evaluation and Feedback.* Using the observation forms done by the participants in field training, a reliability check for each tape will be made. Items will be corrected, and uniformly missed items will be replayed and reviewed. (45 min.)

(7) *Recommendations and Decision Making.* Using one of the participants' corrected observation forms for each of the eight tapes, a composite form will be made on a transparency and the items totaled. The pattern of gaps and strengths will be analyzed. Using the pattern, the participants will then be required to write a recommendation to an advisory board as to what actions should be taken in order to strengthen pupil and teacher learning in gap areas.

One participant will be chosen to act as the person making the recommendations to the board. The other participants will be assigned various roles as board members, such as student teacher, principal, community leader, or director. The board will debate the recommendations and determine the decisions to be made relating to pupil and teacher learning. The following example of an actual memo is the type of decision that can be made. (45 min.)

(8) *Reassessment.* For the final reevaluation the participants will observe themselves going over the entire observation system by viewing video clips of themselves as learners during this program. They will use the observation form to determine the range of instruction that has been provided. (20 min.)

(9) *Anonymous Evaluation of the Session by the Participants.*

Development of Instructional Programs

The institutional profile has now been established in the areas of assessment. Each area of assessment must be matched to an area of instruction. Some missing areas (gap areas) need to be developed. This session will show how to develop and improve existing instructional modules in these areas.

An instructional module must have certain essential components, such as: (1) type of skill or task (knowledge-lecture) to be performed, with the characteristics broken down; (2) demonstration showing alternate ways of performing that task; (3) a one-page analysis of the major elements in the demonstration; (4) an exercise for the participant; (5) video feedback; (6) a retrial; (7) proficiency test with a pass/fail standard establishing that

the teacher learner is ready to perform the taught skill in the field; (8) specific requirements or criteria for competency which will be used by supervisors in the field; and (9) an evaluation of the module by the teacher learners to be made immediately after the instructional session.

The cluster of the institutions' modules constitutes the curriculum of the training program. From this curriculum, a teacher learner profile is developed

(1) *Exercise.* The participants will view capsule modules and rate them according to the nine module components and related criteria. Volunteer participants will do a capsule module on a familiar, everyday subject. These will be rated by the other participants and remediation for missing areas will be developed with the group

(2) *Practical Exposure.* The participants will view two practice video clips of modules in gap areas already tested in the field. (30 min.)

(3) *Development.* The participants will break into two groups and each will make a minimodule in a priority area. They will teach their module to each other and evaluate it for means of improvement. This represents the pilot exercise for developing a full-scale module. (75 min.)

(4) *Practical Experience.* The participants will act as an advisory board in order to designate tasks and necessary work. A framework will be established for maintaining a teacher-learner profile including their students' records from the beginning to the completion of their certification program.

An outline of the instructional sequence necessary for achievement can then be determined on this profile with respect to its specific elements as a curriculum. A timetable for the development of needed elements in the instructional program will then be formulated.

At this point, using all the data from the three training sessions, a new institutional profile will be recognized as a new responsibility. (80 min.)

(5) *Anonymous Evaluation of the Session by the Participants.*

22 Implementation of CBTE - Viewed as a Developmental Process*

Gene E. Hall

Innovation adoption in educational institutions is an activity that is often approached with little indepth calculation and anticipation of potential consequences. It is all too common to observe an air of casualness that is alarming to those repeatedly exposed to change. Often it seems that decision makers decide to adopt one of the recently developed complex innovation bundles as (e.g., individually guided education; competency based teacher education) so casually as they would change a textbook series, which is a change that should not always be taken lightly. Adoption of innovations in educational institutions is not a simple phenomenon or singular event. Adoption of educational innovations is a process that generates diverse and, in many cases, altogether unexpected outcomes. Innovation adoption can consume much of the energy of the people involved and can cost a great deal in terms of resources, dollars, personnel productivity, and, not least of all, time.

In this paper, an attempt will be made to share with CBTE program managers the experiences, theory, and research findings of the CBAM project staff of the Research and Development Center for Teacher Education. This staff has been actively involved in studying the CBTE program development and adoption processes in teacher training institutions around the Nation. Out of this work and experience a model of "innovation adoption" in educational institutions has been developed. Based on this model, measurement procedures and principles for "adoption agents" are being developed and

* The research described herein was conducted under contract with the National Institute of Education. The opinions expressed are those of the author and do not necessarily reflect the position or policy of the National Institute of Education and no endorsement by the National Institute of Education should be inferred.

researched. Each of these areas of work will be briefly explored in this paper. Implications for planned change will also be presented as food for thought for CBTE program managers.

CBAM Project Overview

Researching, planning, and managing innovation adoption in educational institutions are the foci of the work of the Concerns Based Adoption Model Project of the Research and Development Center for Teacher Education of the University of Texas at Austin. In this work, innovation adoption is viewed as a complex process rather than as a singular event. Phases of this process have been identified and described and are being studied. We hypothesize that each and all individuals within an educational institution progress through a series of growth steps in developing their capability to effectively use an innovation. We further hypothesize that adoption agents who assess these developmental stages in their clients improve their ability to select and make personalized interventions, thereby accelerating advancement to effective and extensive use of the innovation within the institution.

A concentrated attempt is being made to conduct empirically based studies and at the same time develop procedures and techniques that have immediate values for adoption agents and program managers. Although much of the early research and development was done in institutions that were developing competency based and personalized teacher education programs, the findings seem to be applicable to other educational institutions and other educational innovations. The research thrusts are targeted toward several fronts:

1. Identification and description of the phases and steps entailed in innovation adoption in educational institutions
2. Development of assessment methods for predicting user system potential for successful adoption of innovations
3. Development of measurement procedures for assessing and diagnosing the developmental stages and levels of individuals and the composite user system
4. Transfer to and immediate application of the research techniques and findings to the activities of adoption agents
5. Using the work and experiences of adoption agents as a source of hypotheses and as a heuristic for learning more about innovation adoption in educational institutions

The remainder of this paper is devoted to brief descriptions of each of these research and development activities.

Institutional Phases in the Adoption Process

Hall (1973), in viewing the history of educational institutions over an extended period of time, has identified what appears to be an ebb and flow in the intensity of innovative activity. Educational institutions seem to oscillate between extended periods of quiet and relatively shorter periods of great activity. The extended periods of quiet have been labeled *periods of*

equilibrium, and the shorter, highly active periods have been named *periods of disequilibrium*. This does not mean to say that during periods of disequilibrium everything is out of control, but rather that the user system as a whole and the individuals within it are experiencing unusual events, problems, imbalance, and increased dynamism. However, to many of the institutions' members a period of disequilibrium may seem to be a time when things are out of control—as if the day-to-day hum of routine were disrupted with the cacophony of builders with jackhammers attempting to build a new order to things. In most cases this building means tearing down some of the old structures, changing work patterns, and meeting new challenges. These changes will be perceived differently by each person.¹

If the attempt at innovation adoption is completely successful, institutions are hypothesized to move through six phases of disequilibrium. These phases are:

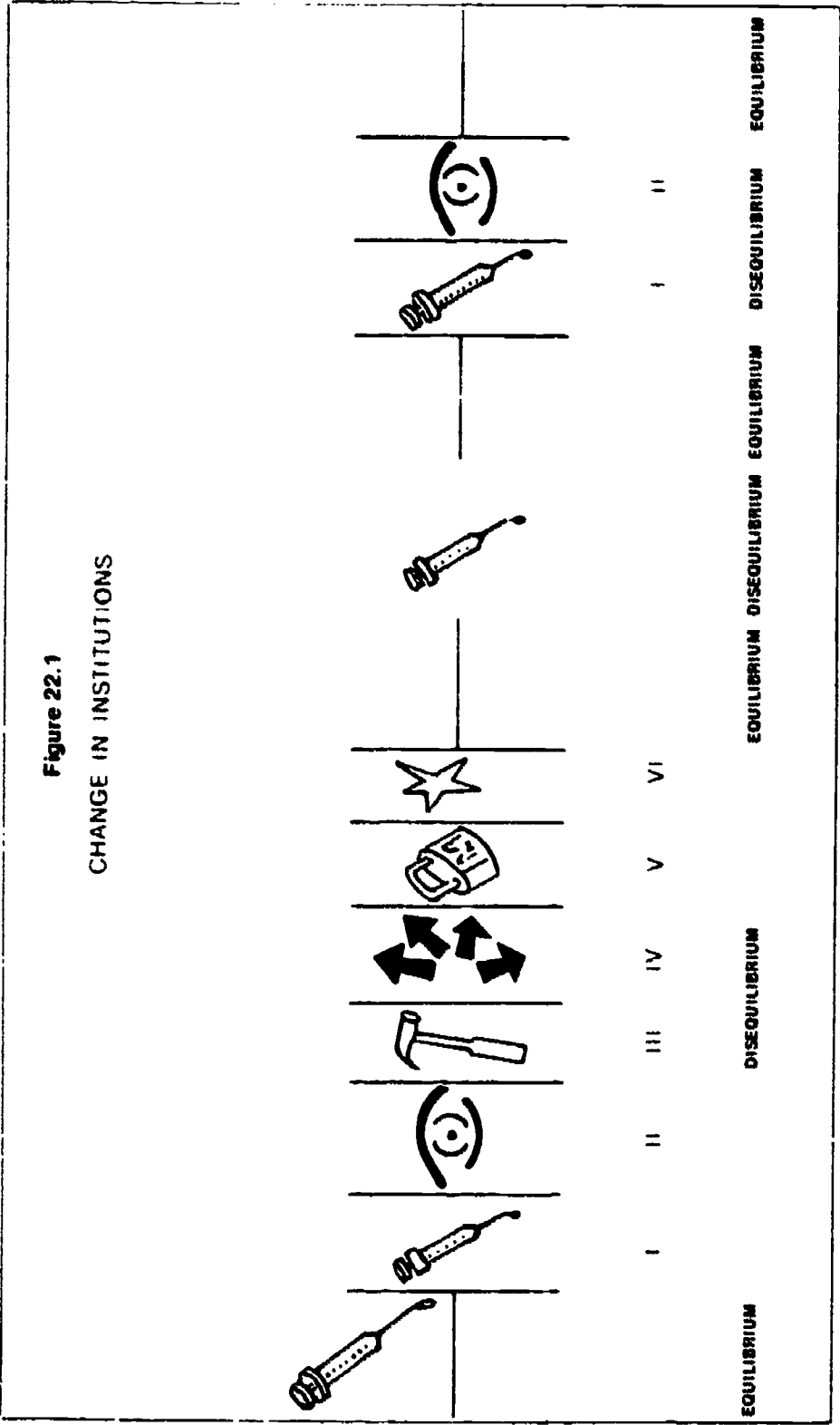
1. *Injection*—the idea of the innovation is introduced to members of the institution
2. *Examination*—the innovation receives study, talk, visits, thought, planning, research, and committee formation
3. *Preparation*—the time following the commitment to try out the innovation when materials and resources are organized and pre-use training occurs
4. *Sampling*—first tryout of the innovation on an experimental basis by part of the user system
5. *Spread*—the spread of trial use of the innovation to all potential users within the user system
6. *Institutionalization*—the innovation comes into regular use by all or nearly all potential users.

As is represented in Figure 22.1, institutions appear to oscillate between periods of equilibrium and periods of disequilibrium. Within any period of disequilibrium, the number of phases and the duration of each phase is dependent on a number of variables, such as the complexity of the innovation, the users' ability to use similar innovations, the users' level of use of previously adopted innovations, the leadership of the adoption agent(s), and a host of other variables. Termination of attempted innovations at one or another phase of adoption is an all too common occurrence.

Predicting Potential for Successful Adoption

Manning (1973) has developed an experimental instrument entitled "The 'Trouble Shooting' Checklist" (TSC) for assessing a teacher training institution's potential for successful adoption of two innovations: instructional modules and a counseling technique called Personal Assessment Feedback. The TSC can be completed by an adoption agent during an initial visit to an institution that is considering adoption of an innovation.⁴

Figure 22.1
CHANGE IN INSTITUTIONS



Through extensive review of the research literature and intensive interviews with successful change agents and analysis of their responses to early forms of the TSC, Manning has developed a set of items that the adoption agent rates for each of the following categories

1. Organizational structure
2. Personality and leadership styles of organization members
3. Communications
4. Level of usage
5. Characteristics of students within the institutions

A series of subfactor and factor scores result which make possible the classification of an educational institution into one of three categories. The institution is classified as being either

1. An ideal institution for successful adoption of innovations
2. A marginally acceptable institution for successful adoption of innovations
3. An unacceptable institution for successful adoption of innovations

Once this classification is made, the TSC manual provides the program manager or adoption agent with information about a likely sequence of events in the adoption of an innovation in that category of institution. The TSC manual also includes information for the adoption agent about the skills that are most likely to be effective given the institution's predicted potential for successful innovation adoption. These guidelines for adoption agent skills are classified under five headings

1. Leadership Style
2. General Cognitive Skills
3. General Communication and Interpersonal Skills
4. Relationships that the Adoption Agent has with the Faculty
5. Relationships that the Adoption Agent has with the Administration

Wallace (1973) has also identified a set of principles for effective change agents and elaborated on these. This complementary set of guidelines needs to be considered in planning for innovation adoption.³

Figure 22.2 is a sample of the TSC Form A for module-adopting institutions. The subcategory is Organization Structure; the adoption agent is asked to select the eight items out of the list that most directly apply to the user system being considered. This procedure is then repeated for the other TSC categories. The items are then assigned a weight and the weights are summed up to yield a subfactor score. In Category A, the items are weighted as follows

1) 0	5) 0	9) 2	13) 2	17) 1	21) 1
2) 1	6) 1	10) 2	14) 2	18) 0	22) 2
3) 1	7) 1	11) 1	15) 2	19) 0	23) 0
4) 2	8) 0	12) 0	16) 2	20) 1	24) 0

The subscale score range has been tentatively assigned as follows:

Score Ranges	0-4	unacceptable
	5-10	marginally acceptable
	11-16	ideal

Research efforts are presently underway to validate these score ranges and to refine the TSC. Interpretation of the data must be carefully weighted by the CBTE program manager, however, the factors and factor scores should provide him with some added information (although tentative) and insight into what the real potential capabilities and inherent risks are in facilitating a given institution's adoption of CBTE.

Figure 22.2

TSC-A*

(FOR MODULE-ADOPTING INSTITUTIONS)

SECTION I

The following TSC categories and items focus on the institution's organizational structure and include characteristics of the faculty and administration as they relate to organizational structure.

CHECK ONLY THE 8 ITEMS THAT MOST APPLY.

Category A: Organization Structure

- _____ 1. The internal change agent working at this institution appears to be incompetent, and his position lacks authority and responsibility
- _____ 2. There is little state level support or leadership
- _____ 3. The group of potential adopters seems to have some communication problems with the larger faculty group.
- _____ 4. There is a small group of adopters which has credibility with a larger faculty group that gives feedback.
- _____ 5. The potential adopters that do exist have serious communication problems with the larger faculty group.
- _____ 6. The internal change agent working at this institution, although quite capable, is not in a position of authority.
- _____ 7. It is not yet clear how large the group of adopters will be.
- _____ 8. The internal political structure is such that the tenured faculty exerts pressure against innovation.
- _____ 9. There is an "intellectual" authority figure in addition to "line-staff" authority
- _____ 10. The organization has a stable structure with fairly well defined roles and established (functional) channels of communication.

* Brad A. Manning, "The Trouble-Shooting Checklist: A Manual to Aid Educational Change Agents in the Prediction of Organizational Change Potential" (Austin: University of Texas, Research and Development Center for Teacher Education, 1973).

tion

- _____ 11. There is no "intellectual" authority figure—only "line-staff" authority.
- _____ 12. The source of power lies outside of the institution.
- _____ 13. The internal change agent working at this institution is in a position of authority and responsibility.
- _____ 14. There is a small group of highly involved adopters who work in close proximity.
- _____ 15. There is a small group of adopters appearing to move faster and more effectively than would a large group of adopters.
- _____ 16. There is a small group of adopters who clearly demonstrate an ability to effectively communicate with a larger faculty group in order to gain their support
- _____ 17. There are a number of potential adopters, but none who are yet fully committed.
- _____ 18. Potential adopters are scattered across campus and do not have daily contact.
- _____ 19. There is a closed organizational structure. (All activities fit into a predetermined structure.)
- _____ 20. There is a strict, hierarchical organization.
- _____ 21. The group of adopters has not yet established credibility with a larger faculty group but clearly shows potential to do so.
- _____ 22. The organization structure includes the following hierarchy of positions: president; provost; dean; and department chairman.
- _____ 23. There are no *committed* adopters or potential adopters identifiable
- _____ 24. Those individuals who have expressed interest in the innovation have low credibility with the rest of the faculty and appear to be locked into their positions.



CATEGORY 1-A SCALE SCORE

Concerns Based Adoption Model (CBAM)

Hall, Wallace, and Dossett (1973) have developed a model to represent the complex process of innovation adoption in educational institutions. The model—the Concerns Based Adoption Model (CBAM)—is an attempt to represent the complex innovation adoption process by clustering the many possible variables into a condensed set of dimensions and interrelationships that can be used for study. A model that is as complex as the process represents would not be of value to either the researcher or the practitioner. The CBAM is based on viewing innovation adoption as a developmental process. The CBAM organizes the innovation adoption process around the concepts of collaborative linkage, individuals' use of the innovation and individuals'

concerns about their use of the innovation. Information about use and concern provide the adoption agent with the basis for selecting personalized assistance strategies.⁴ The basic dimensions of the CBAM are explored in the next few pages.

Description of CBAM

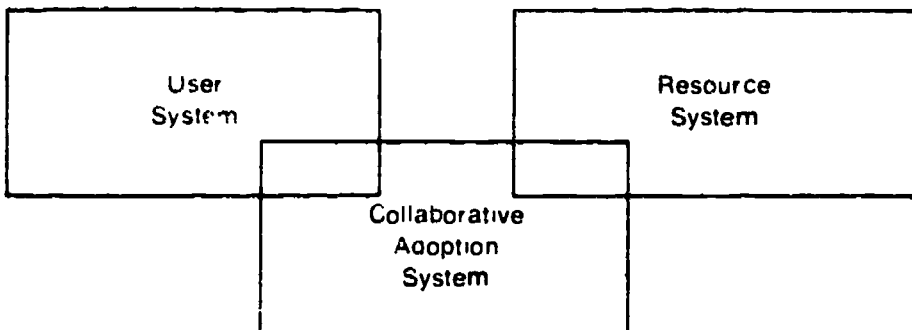
Collaborative Linkage

The CBAM begins with viewing the adopting institution as a *user system* composed of individuals, each of whom has his own set of concerns, problems, skills, agendas, and needs. In combination, these individuals represent the institution and its functioning. When this user system becomes involved in adopting an innovation, a *resource system* that is expert in the use of the innovation is normally available to help it develop its capability. Sometimes the resource system is an individual, sometimes it is located with the user system, usually, however, it is a formal organization outside the user system that forms a linkage with the user system.

Whatever form the resource system takes, for best results with all complex innovations the linkage should be a collaborative one based on mutual openness in communication and a sharing of resources, investments, outcomes, and risks (see Figure 22.3). A one-way association is not likely to survive because the receivers will not sustain a commitment to a joint effort. The CBAM requires that investments be made by both user and resource systems, and that both be able to gain from the collaboration. In most instances a *collaborative linkage* is established to help the user system develop an effective use of the innovation as quickly and as easily as possible. This means that with time the individuals within the user system must become as knowledgeable about the innovation as are the members of the resource system. In addition, each individual in his role—whether it be as an administrator, faculty member, or student—must develop the skills and finesse in using the innovation that will optimize the effects of its use.

Figure 22.3

BASIC REPRESENTATION OF THE SYSTEMS OF THE CONCERNS BASED ADOPTION MODEL



One premise underlying the CBAM is that adoption agents (specialists in the use of the innovation and effective catalysts for facilitating change) work with people in the user system both individually and in groups. As a result, the CBAM at one level focuses on assessing the current state of the individuals within the user system. This assessment then allows the adoption agent to focus his interventions so that they respond to the perceived needs of the individual users and also relate to their levels of use of the innovation at that time. The constructs of the CBAM that make this possible are the two sets of scales

1. Levels of use of the innovation
2. Stages of concern about the innovation

In addition, a third and more specific set of hypotheses has to do with the relationship of stages of concern to levels of use.

Levels of Use of the Innovation

We contend that there are observable differences in how various individuals approach and use an innovation. Specifically, it is hypothesized that there are identifiable, definable, and measurable levels of use of an innovation that range from lack of knowing that the innovation even exists into an active, sophisticated, and highly effective use of it. It is further hypothesized that improved use of the innovation by most individuals is a developmental process. Normally, individuals do not use an innovation for the first time, or even the second time, as efficiently and effectively as do those who have been involved with the innovation through four or five cycles of use. Advanced levels of use are not attained merely by repeated use of the innovation, however. Experience is essential but not sufficient to insure that a given individual will develop the ability to use an innovation effectively.

An oversimplified but helpful illustration of the level-of-use dimension is the process a college instructor goes through when he adopts a new textbook for a course he has taught many times. At first he will use the new book for reference much more than he did the old text. In preparing class presentations and examinations, he will refer to it much more. His assignments are more likely to be specific chapter assignments, and he probably will follow a straightforward progression through the text. His use of the innovation is apt to be mechanical, uneven in flow, and closely related to the text. As this instructor prepares to teach the course a second time using the new text, however, he is likely to select a different arrangement of assignments. This time, he may assign Chapter 4 first and, perhaps, delete Chapter 7 while substituting another reference he thinks will do a better job. In making these changes, he has progressed beyond a mechanical use of the innovation. He has gained the experience and know-how to be more adaptive in his approach, and he more smoothly integrates the use of the text into the rest of his instructional activities.

For research, operational definitions and scale points for the levels-of-use dimension of the CBAM have been developed. To illustrate, Table 22.1 contains sample behaviors found at each level. Two subscales are hypothesized for the levels-of-use dimension. One describes the *knowledge* level of the

user. It hypothesizes that the cognitive level, or amount of information and degree of understanding an individual user has about the innovation, is a developmental progression. Assessment of this set of scale points might take the form of a pencil-paper achievement test. The other set of scale points for level-of-use of the innovation are the *action* scale points. In the CBAM, one hypothesizes that there are observable behavioral differences in how the innovation is actually used and that advancement to the higher levels of use of the innovation is a developmental process. Assessment of the action level of use requires direct observation of the users while they use the innovation.

Table 22.1

**LEVELS OF USE AND TYPICAL BEHAVIORS
FOR EACH LEVEL OF USE OF THE INNOVATION**

Level of Use	Behavioral Indices of Level
Non-Use	No action is being taken to learn about new ideas in the area of the innovation.
Orientation	The user is seeking out information about the innovation.
Initial Training	The user is preparing to use the innovation.
Mechanical	The user is using the innovation in an awkward, poorly coordinated manner.
Independent	The user is doing a good job with the innovation.
Integrated	The user is sharing with others what he is learning about students from using the innovation.
Renewing	The user is seeking out more effective alternatives to his established use of the innovation.

Stages of Concern About the Innovation

A second dimension has to do with the individual user's needs, motivations, problems, and requests as he is becoming expert in using the innovation. In a way that parallels Fuller's (1969) studies of concerns of teachers, individuals are hypothesized to have concerns that relate to their potential or actual use of an innovation. A set of scale points—Stages of Concern About the Innovation—has been defined for this dimension, and it is hypothesized that this dimension is also a developmental progression. That is, when the individuals first approach using an innovation, their concerns will be different from those they will have after they have used it a while. Still higher stages of concern will be expressed with subsequent cycles of use unless one or more developmental processes become blocked or dormant.⁵

BEST COPY AVAILABLE

As with Fuller's theory of concerns of teachers, the CBAM hypothesizes that early concerns are much more self-oriented than are later concerns. Table 22.2 lists Stages of Concern About the Innovation ranging from Unaware to Renewal with typical expressions of concern.

Table 22.2
STAGES OF CONCERN AND TYPICAL EXPRESSIONS
OF CONCERN ABOUT THE INNOVATION

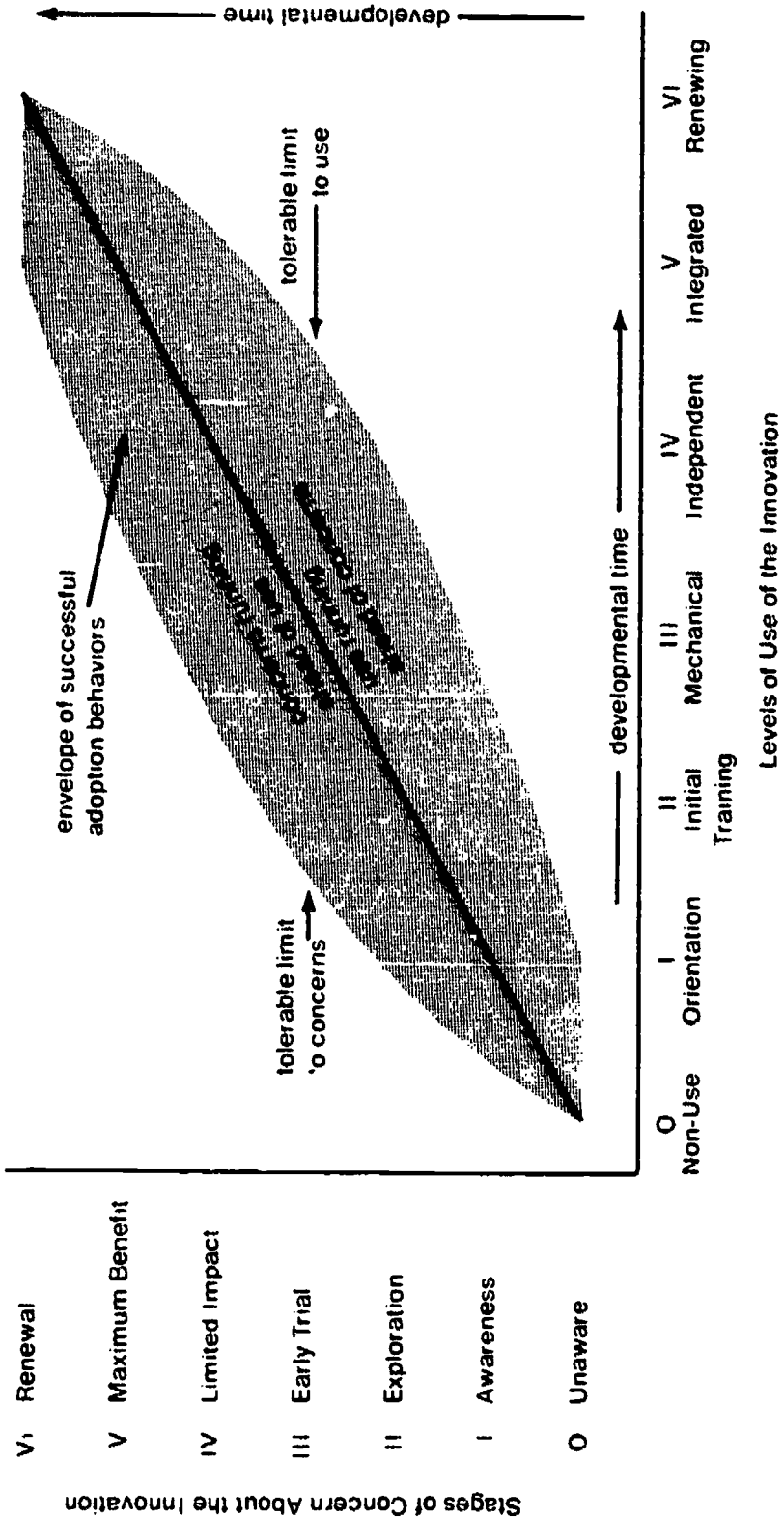
Stage of Concern	Expressions of Concern
Unaware	I don't know anything about it (the innovation).
Awareness	I have heard about the innovation, but I don't know much about it.
Exploration	How much of my time would use of this innovation take?
Early Trial	I seem to be spending all my time in getting material ready for students.
Limited Impact	I can now see how this innovation relates to other things I am doing.
Maximum Benefit	I am concerned about relating the effects of this innovation with what other instructors are doing.
Renewal	I am trying a variation in my use of the innovation that looks like it is going to result in even greater effects.

Relationship Between SoC and LoU

It is hypothesized that concerns are related to use and that it is possible for change agents to infer a great deal about use of the innovation from listening to the user's concerns. This relationship is not always a simple one-to-one correspondence, however. Many of us, for example, have known golfers who "talked a good game" but whose actual play was rather far over par. The alternate imbalance in theory is also possible where the individual's concerns are of a very low level and he has serious doubts about his abilities when, in fact, he has the potential of being outstanding. There are also instances of individuals who "perform over their heads."

An illustration of these relationships using an educational innovation could be schools adopting open concept classrooms. Many communities now have school buildings that are open concept and have reputations for having exciting, innovative programs. When one visits some of these schools, however, he finds that bookcases, chalk boards, easels, and seating are arranged in blocks that serve as traditional self-contained classrooms (low

Figure 22.4
RELATIONSHIPS BETWEEN STAGES OF CONCERN
AND LEVELS OF USE WITH SUCCESSFUL ADOPTION



use). In another school, where all the walls are gone, the pupils are roving aimlessly. No territories have been established; there is excessive confusion; and the climate feels tense. This is a school that is probably early in its use of open-concept classrooms and where the teachers have high stages of concern about sharing their leadership and responsibility for curriculum and about remaining nonauthoritarian. But, in spite of these high concerns, their level of use of open-concept classrooms is low. They are confused and uncertain as a result of perhaps attempting too ambitious a beginning.

With the CBAM, it is hypothesized that there is probably a middle range of relationship between concerns and use where successful advancement or growth is possible, but if an individual's stage of concern and level of use move too far out of correspondence, then adoption of the innovation is in jeopardy. Figure 22.4 is a graphic representation of this set of hypotheses, with the area within the envelope representing the hypothesized safe-growth area.

Extensity

The ultimate criterion in any innovation adoption effort is the extent and quality of use by each user of the innovation within the user system. The level-of-use dimension of the CBAM contains a set of operationally defined scale points that provides behavioral indicators of the quality of use of an innovation by each individual within the user system. Innovations are adopted by user systems composed of many individuals. It is important to have a record of each individual's level of use. Also, a representation of the proportion of individuals within the user system that are using an innovation needs to be made. A descriptive statement that the average user in a school is at a mechanical level of use is not as useful as is a picture of the present level of use that each individual is demonstrating. An *extensity profile* can be constructed to accomplish this. All faculty, administrators, and students can be observed and rated with respect to their levels of use of the innovation. When this information is plotted, the resultant graph represents the individual level of use and the extent of use of the innovation within the user system; at the time the observations were made. By plotting extensity profiles at regular intervals, a visual record can be maintained of the extent and level of use of the innovation. When extensity profiles for different dates are compared, the rate of advancement of innovation use or its arrest can easily be seen. Figure 22.5 is an example of an extensity profile.

Putting It Together

The Concerns Based Adoption Model in its entirety is represented in Figure 22.6. In operation, there is a collaborative linkage established between a user system that is adopting an innovation and a resource system that has expertise with the innovation and facilitating its adoption. In theory, linkage is accomplished via several communication channels that entail systematic

Figure 22.5
EXTENSITY PROFILE FOR UNIVERSITY X
AFTER TWO YEARS EXPERIENCE IN USING INSTRUCTIONAL MODULES

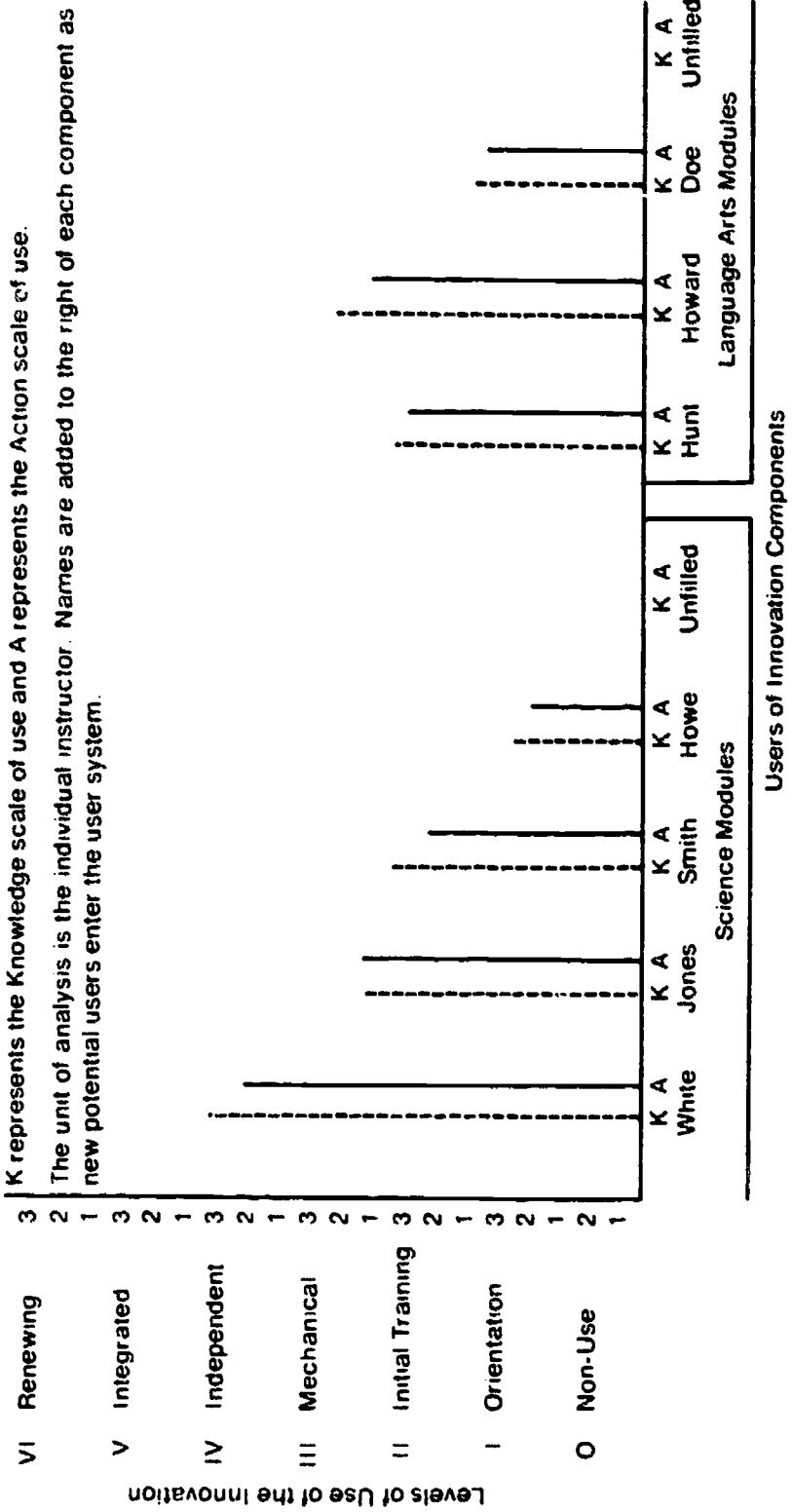
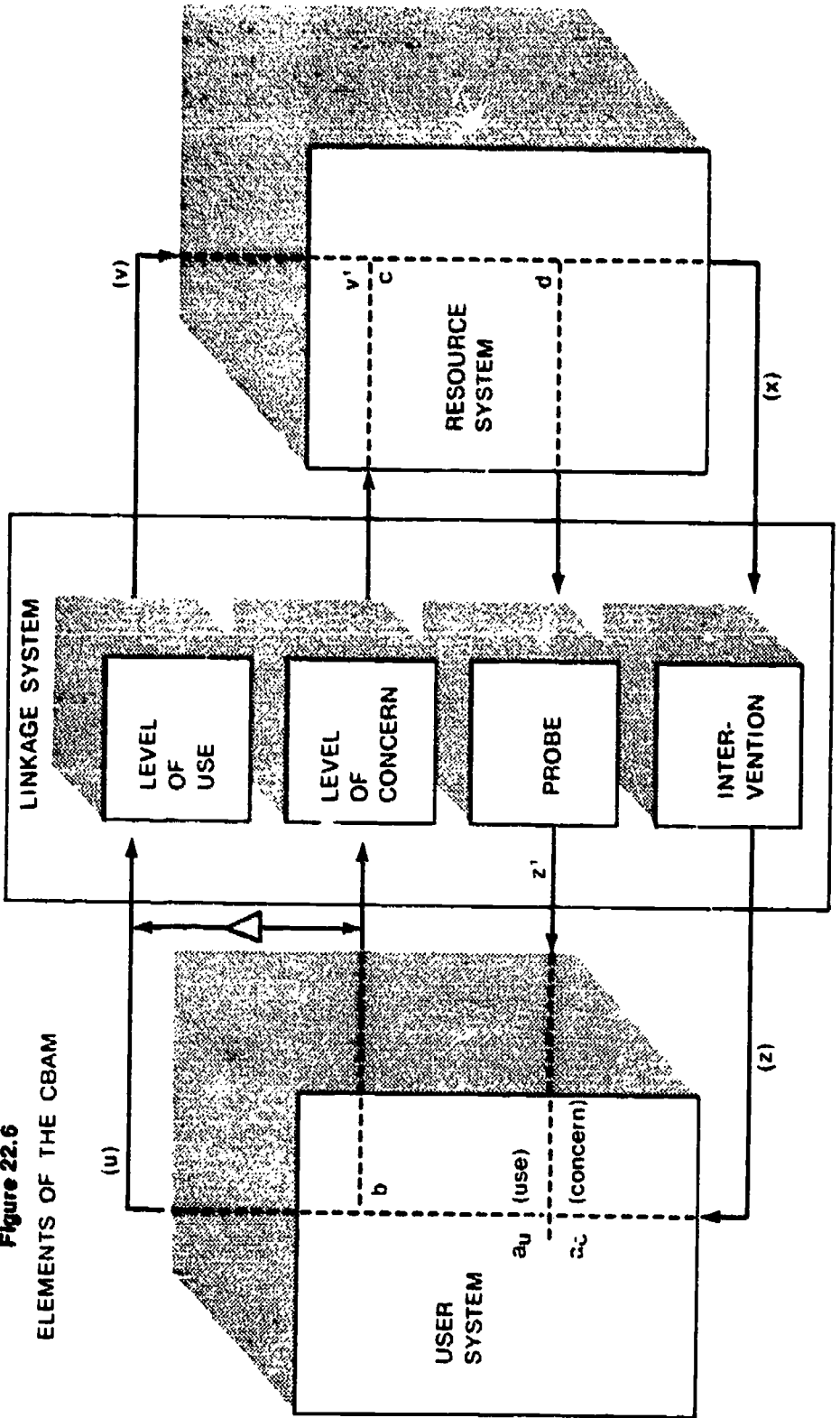


Figure 22.6
ELEMENTS OF THE CBAM



probes of the user system and its personnel to assess each user's stage of concern and level of use about the innovation. Based on this assessment, adoption agents should be better able to select and employ personalized intervention strategies. The selected strategies are targeted toward advancing use of the innovation while, at the same time, resolving the user's concerns or arousing more advanced concerns. Interventions that are targeted in this way are most likely to appear as relevant to the user's concerns and, thereby, are most likely to effect advancement in the level of use of the innovation.

Adoption Strategies

The actual implementation of CBTE can be approached using one or a combination of several different adoption strategies. An adoption strategy is a general game plan or plan of action that is designed to move a user system from first awareness of an innovation to high level and extensive use of the innovation. The adoption strategy selected will greatly influence the rate of spread of use of the innovation, and each strategy has inherent characteristics that can lead to the emergence of certain barriers to successful adoption.

Presently, at least 11 adoption strategies that have been employed extensively can be identified. These are:

1. *The Bootstraps Approach.* An individual within the user system or an entire user system decides to develop or use a new product. Learning how to use it and collecting the necessary resources for using the innovation are carried out by the user system with no outside support or assistance.
2. *The Decree.* An individual in a decision making position within the user system announces that the innovation *will* begin to be used as of a particular time. In many cases, the decision maker is a person who believes in and practices rational decision making. He assumes that everyone else will naturally see the obvious advantages to be gained by using the innovation just because it makes sense. In many cases, few resources and support systems are allocated to the adoption process; it is assumed to be a *fait accompli* with no need for any dragged-out development.
3. *The God-Bless-You Approach.* An innovation representative or a consultant works with the user system for a few hours or as long as one or two days when the innovation is first adopted. He then leaves with the user system left on its own to work out any problems encountered and with the expectation that the innovation will be used effectively.
4. *Intensive Pretraining.* Individuals are introduced to the innovation through a summer workshop, inservice training session, retreat, or short stay at a training center and then begin using the innovation regularly. In some instances, there are occasional followup conferences or meetings during the first year of use.
5. *Sabbatical.* An individual who will use the innovation takes an ex-

- tended leave to spend time at a training center or with an institution already using the innovation. He then can return to his own institution as an advocate of the innovation and facilitate its use.
- 6 **Superstar Strategy** The user system imports one or more recognized experts as full-time members of the system. They may have been involved with the original development of the innovation, or may have gained knowledge and exposure from being with an established user system. They are expected to implant their skills and help the innovation become institutionalized and, at the same time, bring national visibility to the user system.
 - 7 **Experimental Units.** A small group of individuals from within the user system experiments with an innovation on a trial basis. If the innovation is successful, it is implemented throughout the user system. If it fails, a much smaller percentage of resources are lost than would be lost in a blanket adoption.
 8. **Blanket Adoption.** The innovation is adopted for everyone across the user system at the same time. The training that individuals receive varies from none to extensive and prolonged.
 - 9 **Outside Collaboration.** The user system links with an outside resource system for a long-term collaborative relationship. This process allows the user system access to the skills, expertise, and other resources of the resource system and allows the resource system a field basis for evaluating and researching innovation development.
 10. **The Pennsylvania Contingent.** A new group of people is added to the user system, through a change in administration (new superintendent or dean). The new person brings along several associates from his past who work together to bring about change.
 - 11 **Good Time Workshops.** This nonadoption strategy is extensively used in school systems where a set number of inservice days are scheduled. There is no real commitment to an innovation or its adoption. Rather, the goal is to entertain the teachers for the duration of the workshop and have them leaving reporting that it was fun.

If the CBTE program manager consciously plans for and carefully considers the advantages and disadvantages of the adoption strategy selected, he will be better able to handle problems and facilitate CBTE program development and adoption.

Summary

This brief survey of the various research and descriptive works that are underway in the UTR&D/CBAM Project are offered not as final solutions, but as some first attempts to provide program managers involved with the adoption of CBTE programs and other educational innovations with some hooks and handles to use in planning for and managing the innovation adoption process in educational institutions. Having a strong conceptual development and designing effective operational procedures are essential

ingredients for a CBTE program. However, these conceptual ingredients only provide the user system with the package. Putting its contents to use, that is, adopting the innovation, is an equally complicated and essential process.

FOOTNOTES

¹ G. E. Hall, "Phases in the Adoption of Educational Innovations in Teacher Training Institutions" (Austin: University of Texas, Research and Development Center for Teacher Education, 1973).

² B. A. Manning, "The Trouble-Shooting Checklist: A Manual to Aid Educational Change Agents in the Prediction of Organizational Change Potential" (Austin: University of Texas, Research and Development Center for Teacher Education, 1973).

³ R. C. Wallace, Jr., "Each His Own Man: The Role of Adoption Agents in the Implementation of Personalized Teacher Education" (Austin: University of Texas, Research and Development Center for Teacher Education, 1973).

⁴ G. E. Hall, et al., "A Developmental Conceptualization of the Adoption Process within Educational Institutions" (Austin: University of Texas, Research and Development Center for Teacher Education, 1973).

⁵ F. F. Fuller, "Concerns of Teachers: A Developmental Conceptualization," *American Educational Research Journal*, VI, 2 (1969), 207-26.

AUTHORS

Horace E. Aubertine
College of Education
Illinois State University
Normal, Illinois 61761

David C. Berliner
Far West Laboratory
1855 Folsom Street
San Francisco, California 94103

J. Bruce Burke
College of Education
Michigan State University
East Lansing, Michigan 48824

Robert C. Burkhart
Educational Assessment
Systems Corporation
4716 East River Road
Grand Island, New York 14072

Don Davies
44 Union Street
Guilford, Connecticut 06437

C. James Dyer
North Carolina Central University
Durham, North Carolina 27707

Dell Felder
Director, Teacher Corps
College of Education
University of Houston
Houston, Texas 77004

Ned A. Flanders
Far West Laboratory
1855 Folsom Street
San Francisco, California 94103

N. L. Gage
Stanford Center for Research
and Development in Teaching
Stanford University
Stanford, California 94305

Carl A. Grant
School of Education
University of Wisconsin
Madison, Wisconsin 53706

George Gustafson
La Miarda Unified School District
12820 S. Pioneer Boulevard
Norwalk, California 90605

Gene E. Hall
Research and Development Center
for Teacher Education
The University of Texas
Austin, Texas 78712

John H. Hansen
Executive Secretary
Consortium of CBE Centers
P.O. Box 20011
Tallahassee, Florida 32304

BEST COPY AVAILABLE

W Robert Houston
Associate Dean
College of Education
University of Houston
Houston, Texas 77004

Charles E Johnson
College of Education
University of Georgia
Athens, Georgia 30602

Bruce R Joyce
Stanford Center for Research
and Development in Teaching
Stanford University
Stanford, California 94305

David R Krathwohl
Dean, School of Education
Syracuse University
Syracuse, New York 13210

Jorie Mark
U S Office of Education
Washington, D C 20202

Garry L McDaniels
National Institute of Education
815 Brown Building
1200 19th Street, NW
Washington, D C 20208

Frederick J McDonald
Executive Director
National Commission on PBE
Educational Testing Service
Princeton, New Jersey 08540

Donald M Medley
School of Education
University of Virginia
Charlottesville, Virginia 22903

Greta Morine
Far West Laboratory
1855 Folsom Street
San Francisco, California 94103

David A. Potter
Educational Testing Service
Princeton, New Jersey 08540

Marjorie Powell
California Commission
1401 3rd Street
Sacramento, California

Barak Rosenshine
School of Education
University of Illinois
Urbana, Illinois 61801

H. Del Schalock
Oregon College of Education
Monmouth, Oregon 97361

Allen Schmieder
U S Office of Education
Washington, D C 20202

Harvey B Scribner
School of Education
University of Massachusetts
Amherst, Massachusetts

Michael Scriven
Department of Philosophy
University of California
Berkeley, California 94720

Gilbert F Shearron
Chairman, Division of Elementary
Education
University of Georgia
Athens, Georgia 30602

Robert S Soar
Institution for Development
for Human Resources
University of Florida
Gainesville, Florida 32611

Leonard B. Stevens
School of Education
University of Massachusetts
Amherst, Massachusetts

BEST COPY AVAILABLE

Beatrice A. Ward
Far West Laboratory
1855 Folsom Street
San Francisco, California 94103

Wilford A. Weber
College of Education
University of Houston
Houston, Texas 77004

